Alfred Holl

# Data mining and inflectional morphology

**Analogy:**
**a simple strategy of learning**
**Inflectional morphology of the verb:**
**complex relationships within**
**structural systems of natural languages**
**Data mining:**
**discovery of similarities, analogies**

# 0 Motivation and overview: observation

**Mistakes in the language of children**
**Hypercorrect forms of native speakers**
**Mistakes of non-native speakers**
**e.g.**

| | |
|---|---|
| ***he *readed*** | **analogous to *he pleaded*** |
| ***think, thank, thunk*** | **analogous to *sink, sank, sunk*** |
| ****gamlaste*** | **analogous to *lättaste*** |
| ***er *tretet*** | **analogous to *er knetet, betet*** |
| ****getretet*** | **analogous to *geknetet, gebetet*** |
| ***er hat *gelügt*** | **analogous to *er hat gerügt*** |
| ****gestreitet*** | **analogous to *ausgebreitet*** |
| ***ich habe *gewerft*** | **analogous to *ich habe gekauft*** |
| ***du *sehst*** | **analogous to *du stehst*** |
| ***er *frägt*** | **analogous to *er trägt*** |
| ***er *schwomm*** | **analogous to *er klomm*** |

***un *[zwoazo] (oiseau)*** **analogous to *un train – les trains***

# 1 Linguistic analysis

## 1.1 Which strategy leads to these mistakes? Non-reflected use of the assumption of analogy

**The assumption of analogy is only a heuristic method.**

**True assumption of inflectional / morphologic analogy:**

***~ling (~lung, ~lung)****: cling, fling, sling*

***~änka (~änker, ~änkte, ~änkt)****: dränka, kränka, inskränka*

***~reiben (~rieb ge~rieben)****: reiben, schreiben, treiben*

***~acére (~accio ~aci ~acciamo ~acqui ~aciuto)****: giacere, piacere, tacere*

# 1 Linguistic analysis

## 1.1 Which strategy leads to these mistakes? Non-reflected use of the assumption of analogy

**The assumption of analogy is only a heuristic method.**

**False assumption of inflectional / morphologic analogy:**

**~ch**: *hatch, match, watch, fetch* are reg.
*teach (taught taught), catch (caught, caught)*

**~å**: *få, gå, stå, uppnå*
**~inna**: *finna* (oreg.), *hinna* (oreg.), *minnas (-s, -des), tvinna (-ar)*

**~eiben**: to *-reiben* add *einverleiben (-te, -t), bleiben*

**~iegen**:
*biegen (bog gebogen), fliegen, wiegen*
*liegen (lag gelegen)* [*liegt* vs. *tritt* ← *treten (trat getreten)*]
*siegen (siegte gesiegt), kriegen, schmiegen*

**~endere**: not homogeneous due to *pendere, vendere* etc.
**~rendere**: *rendere (rendo rendi rendiamo resi reso), prendere*

# 1.1 Assumption of analogy

**Similarities in the infinitive (lexical base)**
**shall be used as key features (gestalt-psychological term)**
**for the assumption of inflectional / morphologic analogy,**
**that is, for the assumption of the equality of**
**all the essential inflectional / morphologic features, that is,**
**– endings in synthetic tenses**
**– stem alternations in synthetic tenses**
**(Accidental feature: auxiliary verb in analytic tenses**
***essere/avere*, e.g. *essere stato* vs. *avere fatto*,**
***sein/haben*, e.g. *gebogen haben*, *geflogen sein*)**

**Assumption of analogy (German: Analogieschluss, Schluss auf Analogie) is not necessarily true/correct:**
**Starting from the equality of all the key features,**
**the equality of all the essential features is assumed.**

## Possible key features

**– Equality of the infinitive endings (= reverse similarity)**

**– phonetic similarity of the infinitives (Swedish *ä ö* → 2nd conj.)**

**– equal stem alternations within the same *ablaut* class, e.g.**

***binda (band bundit), vinna (vann vunnit), slippa (slapp sluppit)***

***biegen (**biegt **bog gebogen), bieten (**bietet **bot geboten)***

## 1.1.1 Analytic-inductive part

**A lexeme is assumed to be the pattern lexeme
of an entire group of lexemes with the same key features.**

## 1.1.2 Synthetic-deductive part

**Analogy rules:**
**use the known *averbo***
**of a pattern lexeme as paradigm for another lexeme**
**in order to "derive" its (unknown) *averbo*;**

**produce / generate its inflectional forms**
**"in analogy" to the known *averbo*.**

**Averbo: all of the inflectional forms of a lexeme**
**Paradigm: an averbo used as pattern / template**

## 1.2 What is the reason for this strategy?
## Aim: reduction of the learning effort

## 1.2.1 Traditional grammatical description: reduction to essential forms

**Linguists call these essential (or cardinal or fundamental) forms *key forms* or *principal parts* or *stem row*.**

**The linguistic term *"key" form* is not related to the gestalt-psychological term *"key" feature*!**

**The first principal part is the lexical base (in case of verbs the infinitive), the form which is used for quoting a lexeme in a dictionary.**

**Using analogical thinking, the learning effort can easily be reduced when deriving the entire *averbo* of a lexeme from the principal parts.**

**This principle has proved its value in classic grammar books: One learns essential forms (principal parts) instead of complete *averbos*.**

**Examples:**

***teach (teaches taught taught)***
***hinna (hinner hann hunnit)***
***biegen (biegt bog gebogen)***
***rendere (rendo rendi rendiamo resi reso)***

# 1.2.2 Reduction to pattern lexemes

**The learning effort shall be reduced even more.**
**One does no longer want to learn**
**the principal parts of all of the lexemes;**
**learning shall be restricted to**
**the principal parts of pattern lexemes only.**

**This is very easy regarding regular lexemes,**
**but very difficult regarding irregular lexemes.**
**There, it is the reason for the mistakes mentioned.**

**A group of lexemes is called morphologically homogeneous**
**if all of its members are morphologically analogous.**

**There are**
**morphologically homogeneous and inhomogeneous groups**
**of verbs with the same infinitive ending.**

**Knowledge about their distinction is not common:**
**therefore, the assumption of analogy can easily become a trap!**

**Didactic consequences**
**Give learning assistance on the basis of analogy.**
**Help to avoid false assumptions of analogy.**

**There are two competing types of similarity:**
**- reverse similarity**
**- morphological / inflectional similarity**

**Question to a data mining algorithm**
**In which cases do the two types coincide?**

# 1.2.3 Simplification of inflectional morphology: a general tendency

**The mistakes caused by analogical thinking**
**also show the general tendency**
**of analogically spreading "regular" conjugation patterns:**

**In which cases are irregular forms very stable?**
**If they occur very often: many tokens of a type.**

**Otherwise: regular and irregular forms in parallel,**
**partly with semantic or diaphasic differentiation**
**This situation arrives at the replacement by regular forms.**

***learn: learned / learnt; lie: lied / lay, lied / lain***

***fnysa: fnyste / fnös, fnyst; smälta smälte / (smalt), smält / (smultit)***
***sluta: -ar* (komma till slut) / *slöt, slutit* (stänga, dra slutsats)**

***stieben: stiebte / stob***
***weben: webte / wob***
***senden: sendete* (Radio, TV) / *sandte* (Brief)**
***wenden: wendete / wandte***
***triefen: triefte / troff***
***schaffen: schaffte* (fertigbringen) / *schuf* (erschaffen)**
***pflegen: pflegte* (gängig) / *pflog* (obsolet)**
***glimmen: glimmte / glomm***

***concedere: concedei, -etti / concessi, condeduto / concesso***
***perdere: perdei, -etti / persi, perduto / perso***
***partire: partisco / parto* (transitive / intransitive)**

**Only regular conjugation classes (partly with phonotactic, phonetic or orthographic particularities) are productive, that is, new verbs always conjugate regularly, except for fun.**

# 1.3 Motivation for the use of reverse similarity as a key feature 1

**1 Sometimes, the heuristic leads to a correct result.**

**2 A basic verb (*simplex*) and its compound / prefixed verbs have reversely similar infinitives and in the majority of cases the same conjugation.**

**3 Phonotactically, phonetically and orthographically caused specialties of regular verbs can be seen at the infinitive ending. The corresponding inflection subtypes are productive.**

**consonant grapheme doubling (*digging*)**
***e*-insertion in the 3rd singular of present tense for English verbs ending in *–ch, –sh, –Co, –ss, –x, –Cz* (*wishes*)**
***e*-deletion (*aging*)**
***ie-y* change (*dying*)**
***y-ie* change (*studies*)**

**e.g. *samla* 1st conj.; 2nd conj. phonotactically impossible**

***e*-Einschübe bei Verben auf *-den, -ten, -Cmen* (*C* ungleich *h, l, m, r*; alle homogen), *-Cnen* (nur *-fnen, -gnen, -chnen* homogen)**
***e*-Elision im Ind.Prs.1.Sg. bei Verben auf *-eln* und *-ern* (beide homogen)**
***s*-Elision im Ind.Prs.2.Sg. bei Verben auf *-sen, -ssen, -ßen, -xen* (homogen), *-zen*.**

***h*-insertion: *~care, ~gare***
***i*-deletion: *~giare***

# 1.3 Motivation for the use of reverse similarity as a key feature 2

**4 General tendency to simplify inflectional systems (see 1.2.3)**

**5 Only regular inflection types are productive:
new lexemes (neologisms) will never be irregular.
The inflection type is assigned to new lexemes
via reverse similarity to existing lexemes.**

**6 In Latin and the Romance languages,
the assignment to conjugation classes
is based on the infinitive ending.**

**7 The use of end rhymes in poems, children's rhymes,
spoonerisms (Schüttelreime)
attributes a considerable importance to reverse similarity,
while this use is not immediately related to morphology.**

# 2 Goal of morphological data mining: linguistically interesting clusters

## 2.1 Homogeneous clusters (homC) 1

**Mathematically connected set (no gaps or interruptions)
in a reversely ordered sequence
of morphologically analogous lexemes,
if irregular inflection type, not productive
if regular inflection type, productive (new words)**

**1 Many-lexeme homogeneous clusters,
without grapheme type characters and
without word delimitation**

**Example: *~ling (~lung, ~lung)***
**All verbs whose infinitives end in *–ling*.**
**All of the verbs have the same irregularities.**

**Elements:**
***cling (clung, clung); fling (flung, flung); sling (slung, slung)***

**2 Many-lexeme homogeneous clusters (precisely: sets of homogeneous clusters) with grapheme type characters (*C, V*) and without word delimitation**

**Example: *~Cz (~Czed, ~Czed)***
**All verbs whose infinitives end in a consonant grapheme plus *z*.**
**All of the verbs are regular with *e*-insertion in the 3rd person singular of present tense whose ending is extended to *–es*.**

**Elements: *jazz (jazzed, jazzed)* etc.**

# 2.1 Homogeneous clusters (homC) 2

## 3 Many-lexeme homogeneous clusters (precisely: sets of homogeneous clusters) with grapheme type characters and with explicit word delimitation

**Example: *#CCVg (#CCVgged, #CCVgged)***
**All basic verbs whose infinitives have the form**
**"two consonant graphemes plus vowel grapheme plus *g*"**
**and their prefixed verbs.**
**All of the verbs are regular**
**with consonant grapheme doubling**
**in past tense, past participle and gerund.**

**Elements: *drag (dragged, dragged)* etc.**

**Remark**
**Clusters with grapheme type characters (*C, V*):**
**– abbreviations for sets of clusters**
**– mathematically non-connected**

## 4 One-lexeme homogeneous clusters with implicit word delimitation

**They consist of one single basic verb and its prefixed verbs.**
**They have to be dealt with from a formal perspective**
**although they are not interesting from a linguistic one.**

**Examples: *dig (dug, dug); bring (brought, brought)* etc.**

# 2.2 Basic clusters (basC) 1

**Inhomogeneous clusters**
**– most of the verbs have the same morphological properties**
**– threshold percentage of the majority: around 70%**
**– often correspond to traditional inflection classes**
**– if irregular inflection type, not productive (see Italian)**
**– if regular inflection type, productive (new words)**

## 1 Basic clusters without grapheme type characters and without word delimitation

**Example: *~ch (~ched, ~ched)***
**All verbs with infinitive ending *–ch*.**
**Most of the verbs are regular with *e*-insertion in the 3rd person singular of present tense whose ending is extended to *–es*.**

**Elements: *reach (reached, reached)* etc.**

**Exceptions: *teach (taught, taught); catch (caught, caught)***

## 2 Basic clusters (precisely: sets of basic clusters) with grapheme type characters (*C, V*) and without word delimitation

**Example: *~C (~Ced, ~Ced)***
**All verbs whose infinitives end in a consonant grapheme.**
**Most of the verbs are regular.**

**Elements: *look (looked, looked)* etc.**

**Exceptions: *bring (brought, brought)* etc.**
**There are many exceptions, but statistically only a few as the cluster comprises all of the innumerable regular verbs ending in a consonant grapheme.**

# 2.2 Basic clusters (basC) 2

## 3 Basic clusters (precisely: sets of basic clusters) with grapheme type characters and with explicit word delimitation

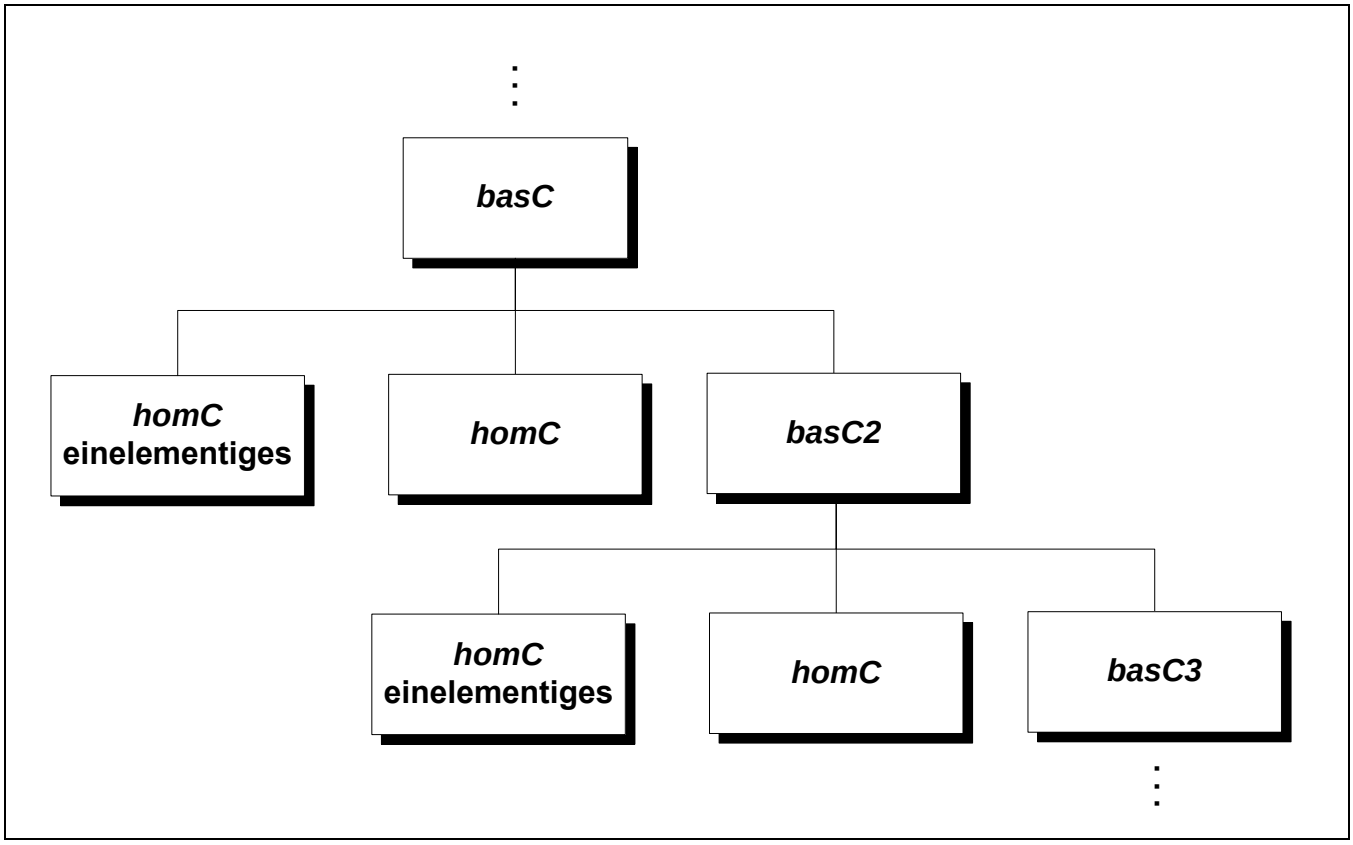
**Example: *#CVg (#CVgged, #CVgged)***
**All basic verbs whose infinitives have the form**
**“consonant grapheme plus vowel grapheme plus *g*”**
**and their prefixed verbs.**
**Most of the verbs are regular**
**with consonant grapheme doubling**
**in past tense, past participle and gerund.**

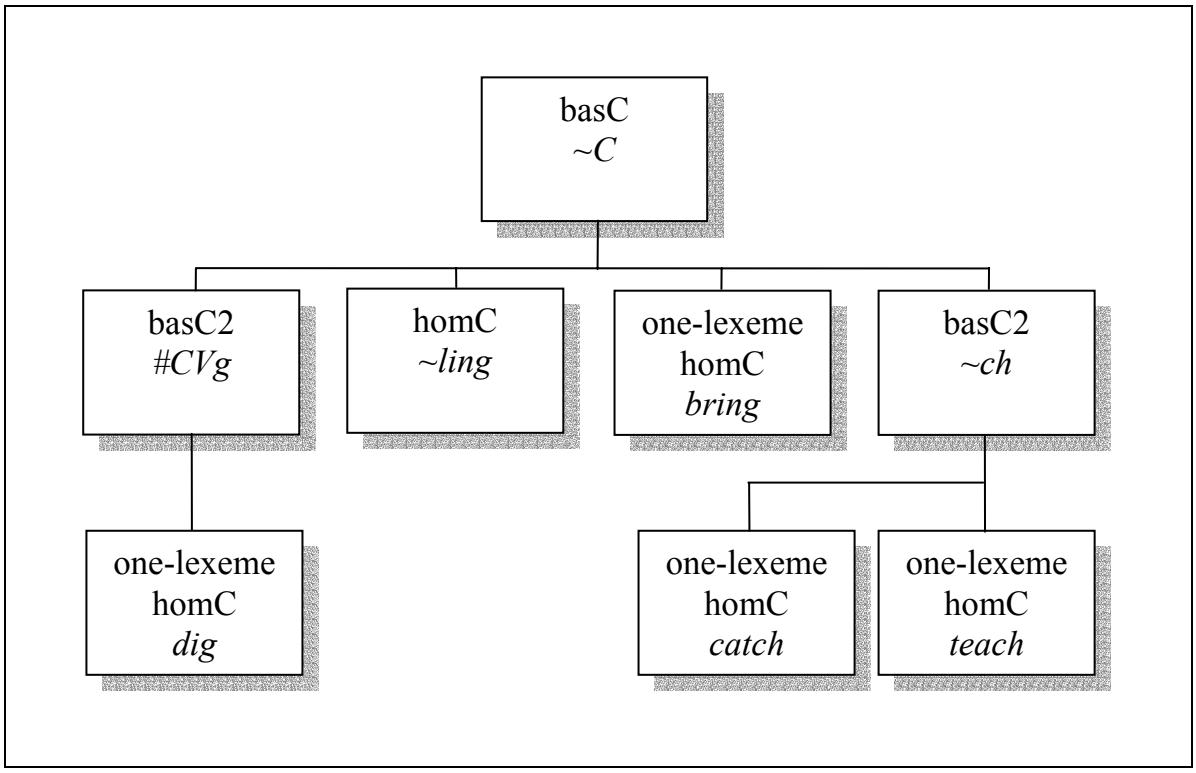
**Elements: *jig (jigged, jigged)* etc.**

**Exception: *dig (dug, dug)***

**Remark**
**Clusters with grapheme type characters (*C, V*):**
**– abbreviations for sets of clusters**
**– mathematically non-connected**

# 2.3 Cluster trees




**Cluster trees: general and a part of the English verbs**

# 2.3 Well-defined cluster trees: cluster hierarchies

**It must be possible**
**to unambiguously assign a search lexeme to its best-fit cluster.**

**Two sets of lexemes belonging to two clusters must be**
**1. either disjoint**
**2. or in a subset-superset relation.**

**The lexical base of a parent cluster always contains**
**fewer or as many characters than the ones of its child clusters.**

**In the case of an equal number of characters,**
**the parent cluster contains more grapheme type characters.**

**That is, regarding lexical bases (alphabetic properties) as conditions:**
**the condition of a parent cluster is always weaker**
**than the conditions of its child clusters.**

# 3 Data mining

## 3.1 The data mining process

### Pre-processing

**Representation of linguistic objects**
**- graphemic, phonemic, phonetic**
**Modified orthographic conventions**
**- prefix treatment, marking of inflectional variants**
**Gathering of linguistic data**
**- reverse dictionaries, verb lists**
**Delimitations**
**- in comparison to syntax and lexicon**
**Principal parts (key forms) and inflection types**
**- selection and definition**
**Derivation rules and exceptions**
**- construction of rules, list of not derivable forms**

### Processing

**Execution of the data mining algorithm**
**independent of**
**the examined language-part-of-speech combination**
**in a inflecting or agglutinating language**
**Top-down cluster analysis**

### Post-processing

**Interpretation, evaluation of the data mining result**
**Preparation for use, typographic marking**
**Basic clusters, reductions**
**Homogeneous clusters**
**Cluster tree**

# 3.2 Data mining algorithm 1

Set *counter_lexeme_ending_length* to *0*

**Outer loop**
Do
while there are lexemes with the entry *NULL* in column *lexeme_cluster*

**Body of the outer loop**
Increase *counter_lexeme_ending_length* by *1*

**Inner loop**
Do while there are data records with
*lexeme_ending_length* = *counter_lexeme_ending_length* - *1* and
*lexeme_cluster* = *NULL*

**Body of the inner loop**
**1. Find reference lexeme**
Take next lexeme with
*lexeme_ending_length* = *counter_lexeme_ending_length* - *1* and
*lexeme_cluster* = *NULL* as reference lexeme

Get current lexeme ending (*counter_lexeme_ending_length*) and inflection type of the reference lexeme

**2. Comparison loop: find numbers of comparison**
*equiv_lexemes* = number of lexemes with the same current lexeme ending and same inflection type as the reference lexeme

*all_lexemes* = number of lexemes with the same current lexeme ending as the reference lexeme irrespective of the inflection type

**3. Evaluate comparison**

*equiv_lexemes* = *all_lexemes*

| true | false |
|---|---|
| <homogeneous cluster found><br>**Evaluation loop**<br>Change attribute values of the corresponding lexemes:<br>*lexeme_ending_length* = *counter_lexeme_ending_length*<br>*lexeme_cluster* =<br>current lexeme ending | <no cluster found><br>**Evaluation loop**<br>Change attribute values of the corresponding lexemes:<br>*lexeme_ending_length* = *counter_lexeme_ending_length* |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---:|:---:|:---:|:---:|
| säen | rglm | 0 | NULL |
| haben | a-a-a | 0 | NULL |
| inne-haben | a-a-a | 0 | NULL |
| vor-haben | a-a-a | 0 | NULL |
| handhaben | rglm | 0 | NULL |
| graben | a:-ä-u:-a: | 0 | NULL |
| um-graben | a:-ä-u:-a: | 0 | NULL |
| beben | rglm | 0 | NULL |
| geben | e:-i-a:-e: | 0 | NULL |
| be-geben | e:-i-a:-e: | 0 | NULL |
| auf-geben | e:-i-a:-e: | 0 | NULL |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---:|:---:|:---:|:---:|
| s**äen** | rglm | 3 | äen |
| ha**ben** | a-a-a | 3 | NULL |
| inne-ha**ben** | a-a-a | 3 | NULL |
| vor-ha**ben** | a-a-a | 3 | NULL |
| handha**ben** | rglm | 3 | NULL |
| gra**ben** | a-ä-u-a | 3 | NULL |
| um-gra**ben** | a-ä-u-a | 3 | NULL |
| be**ben** | rglm | 3 | NULL |
| ge**ben** | e-i-a-e | 3 | NULL |
| be-ge**ben** | e-i-a-e | 3 | NULL |
| auf-ge**ben** | e-i-a-e | 3 | NULL |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---:|:---:|:---:|:---:|
| s**äen** | rglm | 3 | äen |
| h**aben** | a-a-a | 4 | NULL |
| inne-h**aben** | a-a-a | 4 | NULL |
| vor-h**aben** | a-a-a | 4 | NULL |
| handh**aben** | rglm | 4 | NULL |
| gr**aben** | a-ä-u-a | 4 | NULL |
| um-gr**aben** | a-ä-u-a | 4 | NULL |
| b**eben** | rglm | 4 | NULL |
| g**eben** | e-i-a-e | 4 | NULL |
| be-g**eben** | e-i-a-e | 4 | NULL |
| auf-g**eben** | e-i-a-e | 4 | NULL |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---|---|---|---|
| s**äen** | rglm | 3 | äen |
| **haben** | a-a-a | 5 | NULL |
| inne-**haben** | a-a-a | 5 | NULL |
| vor-**haben** | a-a-a | 5 | NULL |
| hand**haben** | rglm | 5 | NULL |
| g**raben** | a-ä-u-a | 5 | raben |
| um-g**raben** | a-ä-u-a | 5 | raben |
| b**eben** | rglm | 4 | NULL |
| g**eben** | e-i-a-e | 4 | NULL |
| be-g**eben** | e-i-a-e | 4 | NULL |
| auf-g**eben** | e-i-a-e | 4 | NULL |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---|---|---|---|
| s**äen** | rglm | 3 | äen |
| **haben** | a-a-a | 5 | NULL |
| inne-**haben** | a-a-a | 5 | NULL |
| vor-**haben** | a-a-a | 5 | NULL |
| hand**haben** | rglm | 5 | NULL |
| g**raben** | a-ä-u-a | 5 | raben |
| um-g**raben** | a-ä-u-a | 5 | raben |
| **beben** | rglm | 5 | beben |
| g**eben** | e-i-a-e | 4 | NULL |
| be-g**eben** | e-i-a-e | 4 | NULL |
| auf-g**eben** | e-i-a-e | 4 | NULL |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---|---|---|---|
| s**äen** | rglm | 3 | äen |
| **haben** | a-a-a | 5 | NULL |
| inne-**haben** | a-a-a | 5 | NULL |
| vor-**haben** | a-a-a | 5 | NULL |
| hand**haben** | rglm | 5 | NULL |
| g**raben** | a-ä-u-a | 5 | raben |
| um-g**raben** | a-ä-u-a | 5 | raben |
| **beben** | rglm | 5 | beben |
| **geben** | e-i-a-e | 5 | geben |
| be-**geben** | e-i-a-e | 5 | geben |
| auf-**geben** | e-i-a-e | 5 | geben |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---|---|---|---|
| s**äen** | rglm | 3 | äen |
| **haben** | a-a-a | 6 | haben |
| inne-**haben** | a-a-a | 5 | NULL |
| vor-**haben** | a-a-a | 5 | NULL |
| hand**haben** | rglm | 5 | NULL |
| g**raben** | a-ä-u-a | 5 | raben |
| um-g**raben** | a-ä-u-a | 5 | raben |
| **beben** | rglm | 5 | beben |
| **geben** | e-i-a-e | 5 | geben |
| be-**geben** | e-i-a-e | 5 | geben |
| auf-**geben** | e-i-a-e | 5 | geben |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---|---|---|---|
| s**äen** | rglm | 3 | äen |
| **haben** | a-a-a | 6 | haben |
| inne-**haben** | a-a-a | 6 | -haben |
| vor-**haben** | a-a-a | 6 | -haben |
| hand**haben** | rglm | 5 | NULL |
| g**raben** | a-ä-u-a | 5 | raben |
| um-g**raben** | a-ä-u-a | 5 | raben |
| **beben** | rglm | 5 | beben |
| **geben** | e-i-a-e | 5 | geben |
| be-**geben** | e-i-a-e | 5 | geben |
| auf-**geben** | e-i-a-e | 5 | geben |

| Lexical base (infinitive) | Inflection type | Lexeme ending length | Lexeme ending |
|---|---|---|---|
| s**äen** | rglm | 3 | äen |
| **haben** | a-a-a | 6 | haben |
| inne-**haben** | a-a-a | 6 | -haben |
| vor-**haben** | a-a-a | 6 | -haben |
| hand**haben** | rglm | 6 | dhaben |
| g**raben** | a-ä-u-a | 5 | raben |
| um-g**raben** | a-ä-u-a | 5 | raben |
| **beben** | rglm | 5 | beben |
| **geben** | e-i-a-e | 5 | geben |
| be-**geben** | e-i-a-e | 5 | geben |
| auf-**geben** | e-i-a-e | 5 | geben |

# 3.2 Data mining algorithm 2

**Outer loop**
For each lexeme in reverse ascending order with *lexeme_cluster* == *NULL*

**Body of the outer loop**
*ref_lexeme* := *lexeme*

*ref_lexeme_length* := number of letters of *ref_lexeme*

*counter_lexeme_ending_length* := *0*

**Inner loop**
while *lexeme_cluster* == *NULL*

**Body of the inner loop**
1. Increase *counter_lexeme_ending_length* by *1*

*ref_lexeme_ending* :=
ending(*ref_lexeme*, *counter_lexeme_ending_length*)

*equivalen_lexemes* := *0* (number of compare lexemes
with equal inflection type as the reference lexeme)

*non_equivalent_lexemes* := *0* (number of compare lexemes
with another inflection type than the reference lexeme)

**2. Comparison loop**
while there is a compare lexeme with
*compare_lexeme_ending* == *ref_lexeme_ending*

*compare_lexeme* := next lexeme in reverse ascending order

*compare_lexeme_ending* :=
ending(*compare_lexeme*, *counter_lexeme_ending_length*)

**Comparison**

*ref_lexeme_infl_type* == *compare_lexeme_infl_type*

| true | false |
|---|---|
| *equivalent_lexemes* := *equivalent_lexemes* + *1* | *non_equivalent_lexemes* := *non_equivalent_lexemes* + *1* |

**Case distinction:**
4 cases (see comment)

# 3.2 Data mining algorithm 2

Top-down cluster analysis strategy (divisive method)

The database established in the pre-processing phase has to be sorted reversely using the column “analytic lexical base”.

The column “lexeme cluster” is initialized with NULLs. It is meant to be filled with the name of the cluster found by the algorithm.

The core of the algorithm (Fig.) consists of two nested loops:
step by step, the outer loop processes all of the lexemes;
step by step, the inner loop processes all ending lengths.

In detail, a general step of the algorithm runs as follows.

*n*-th step of the outer loop:

The next lexeme is picked out which is not assigned to a cluster, that is, which has the initial value in the column “lexeme cluster” (*lexeme_cluster == NULL*). This lexeme is called **reference lexeme** (variable *ref_lexeme*).

*m*-th step of the inner loop:

1. The variable *Counter_lexeme_ending_length* is increased by *1* and thus set to *m*. The algorithm gets the reference lexeme’s ending (*ref_lexeme_ending*, depending on the current value of *counter_lexeme_ending_length*) and its inflection type (**reference inflection type**). The algorithm is now going to compare the next lexemes with the reference lexeme. Therefore, the variables of the following comparison loop are initialized with *0*.

2. In a third loop (comparison loop), all of the lexemes which have the same *n*-digit ending (the same *n* trailing letters in the lexical base), are examined (**compare lexemes**). All of the compare lexemes with the reference inflection type are counted in the variable *equiv_lexemes*. Those with a different inflection type are counted in the variable *non_equiv_lexemes*.

3. Depending on the values of the variables *equiv_lexemes, non_equiv_lexemes* and *counter_lexeme_ending_length*, four cases are distinguished (Fig.).

# 3.2 Data mining algorithm 2

| | equivalent_ lexemes | non_equivalent_ lexemes | counter_lexeme_ ending_length |
|---|---|---|---|
| Case 1 | 0 | 0 | — |
| Case 2 | — | >= 1 | = ref_lexeme_length |
| Case 3 | >= 1 | 0 | — |
| Case 4 | — | >= 1 | < ref_lexeme_length |

**Cases after counting**
**equivalent and non-equivalent compare lexemes**

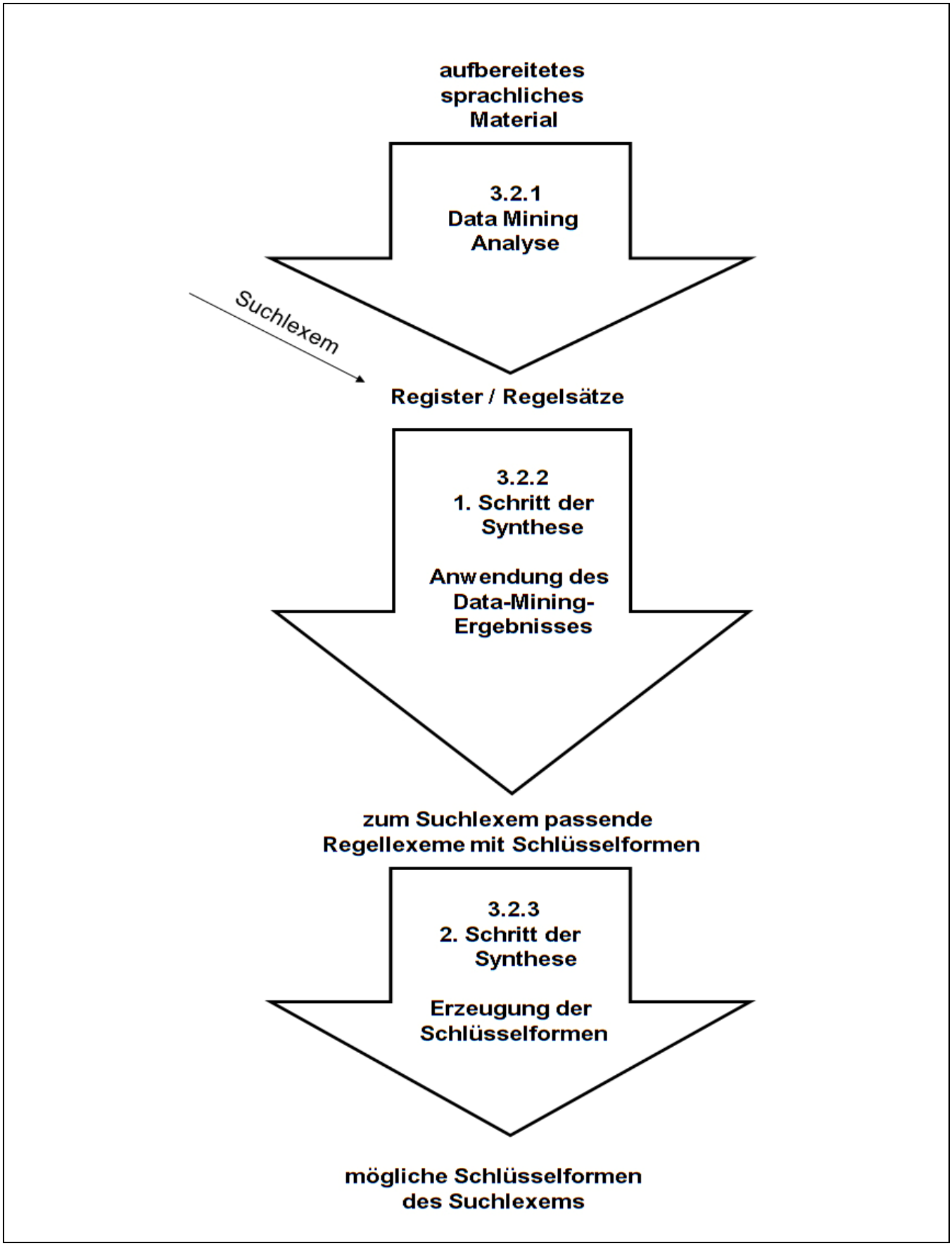
**Case 1 (there are no compare lexemes)**
**one-lexeme cluster consisting of the reference lexeme only**

**Case 2 (there are 1 or more compare lexemes with another inflection type than the reference inflection type, and the current ending length is already equal to the number of letters of the reference lexeme)**
**one-lexeme cluster consisting of the reference lexeme only**

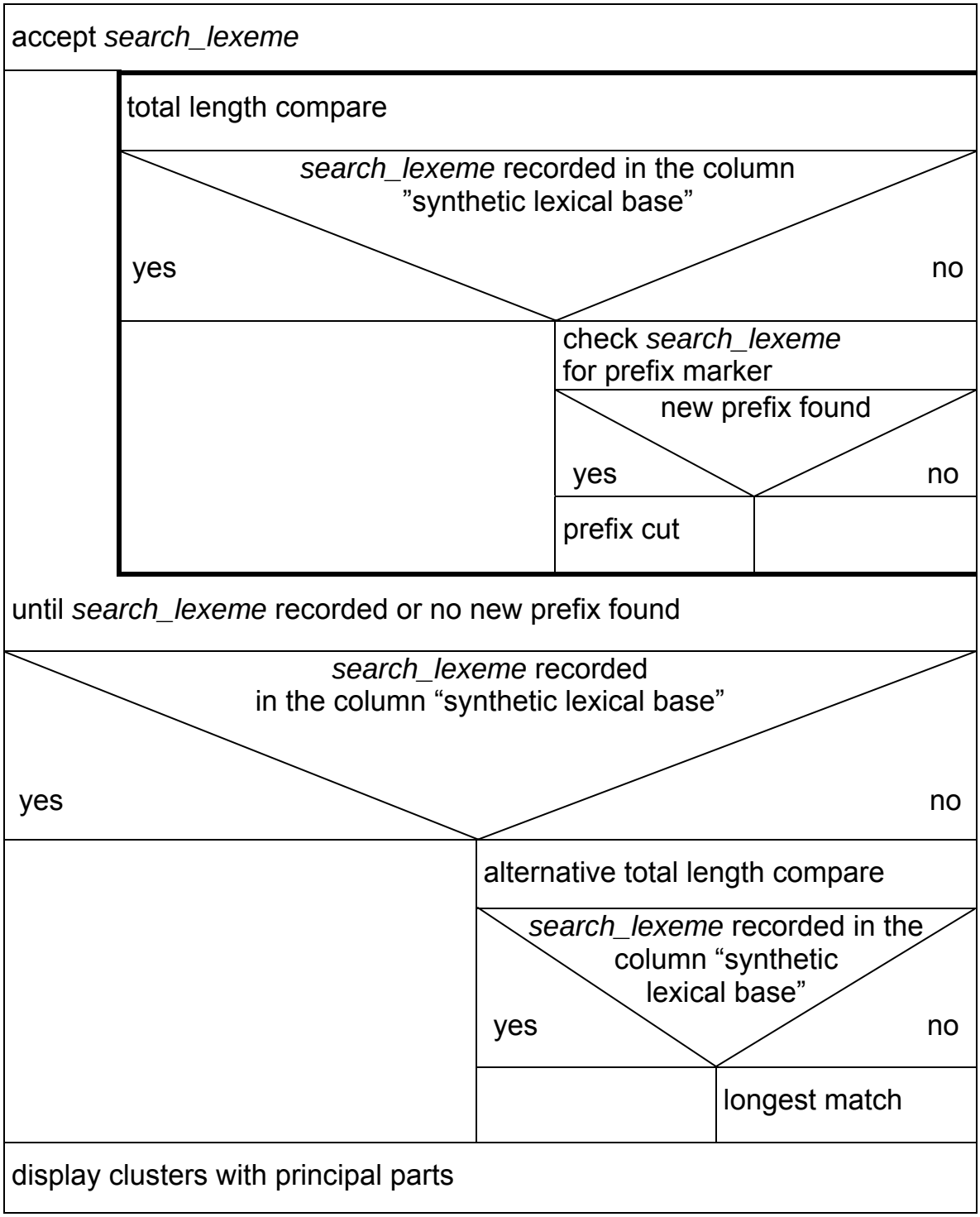
**Case 3 (all of the compare lexemes have the reference inflection type)**
**many-lexeme cluster consisting of the reference lexeme and at least than 1 more lexeme: many-lexeme homogeneous cluster**

**Case 4 (there are 1 or more compare lexemes with another inflection type than the reference inflection type, and the current ending length is still less than the number of letters of the reference lexeme)**
**no cluster**
**The inner loop continues with an increase of the current ending length (*counter_lexeme_ending_length*) by *1*.**

# 4 Search algorithm: the use of the data mining result



# 4 Search algorithm: the use of the data mining result



# 4 Search algorithm:
# the use of the data mining result

**Prefix cut cuts a prefix off the search lexeme.**

**Total length compare finds 1 cluster type:**
**one-lexeme homC.**

**Alternative total length compare creates**
**search alternatives 4 to 0,**
**the *n*-th by replacing normal by grapheme type characters and conserving *n* trailing normal characters.**

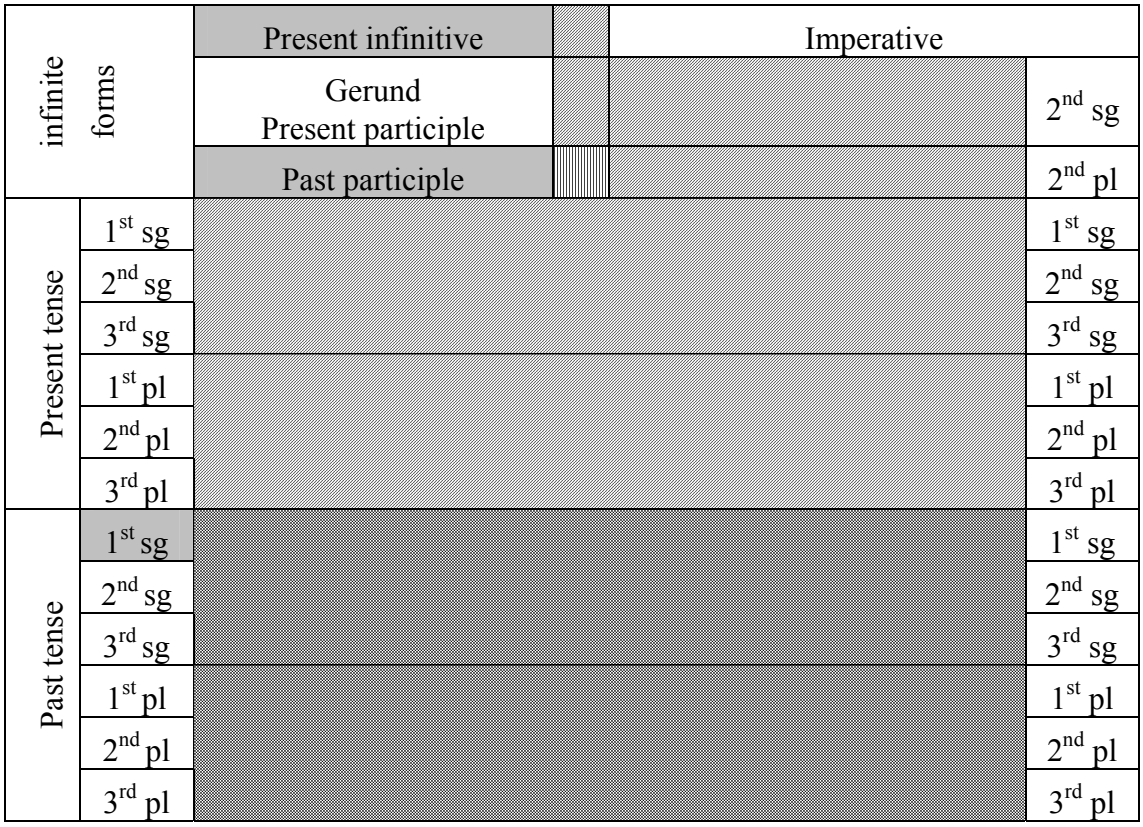
**Alternative total length compare finds 2 cluster types:**
**homC and basC**
**with word delimitation and**
**grapheme type characters.**

**Longest match reduces**
**the search lexeme and its search alternatives**
**from the left by one character after the other.**

**Longest match finds 4 cluster types:**
**homC and basC**
**without word delimitation,**
**no matter whether with or without grapheme type characters.**

# 5.E The situation of the English verb

## 5.E.1 Principal parts




**inflection forms derived from the infinitive**

**inflection forms derived from past tense 1st sg**

**inflection forms derived from past participle**

**Stem distribution of the English verb**
**(Holl / Maroldo / Urban 2007, 94)**

# 5.E.2 Homogeneous clusters

| Infl. type | Lex. base | Examples |
|---|---|---|
| reg / CC | #CVb | rob |
| reg / CC | #CCVb | crab |
| reg / CC | ~CVc | frolic |
| **i-i-i / D** | **~ild** | **gild, build** |
| reg / CC | #CCVg | drag |
| **i-u-u / 0** | **~ling** | **cling, fling, sling** |
| reg / +e | ~sh | fish (fishes) |
| reg / CC | #(C)CVk | trek |
| reg / CC | #CVl | ex_cel, gel, en_rol, ex_tol, an_nul |
| reg / CC | #CCVl | di_stil |
| reg / CC | #CVm | rim |
| reg | ~oo | boo (booing, boos, booed) |
| reg / CC | #CCCVp | strip |
| **ee-e-e / D** | **~weep** | **weep, sweep** |
| **ea-o-o / n** | **~wear** | **wear, swear** |
| reg / +e | ~ss | kiss (kisses) |
| reg / +e | ~x | fix (fixes) |
| reg / CC +e | #(C)CVz | quiz (quizzing, quizzes) |
| reg / +e | ~Cz | waltz (waltzes), buzz (buzzes) |

**homC without examples not mentioned, e.g. #CCCVb**

# 5.E.3 Inhomogeneous clusters

***~ch****: hatch, match, watch, fetch* are reg.
*teach (taught taught), catch (caught, caught)*

**etc. (e.g. see cluster tree)**

# 5.E.4 Basic clusters

| Cluster information | Basic cluster: morphological property (infl. type) | Basic cluster: alphabetic property (lexical base) | Examples | Exceptions |
|---|---|---|---|---|
| basC | reg | ~C | look | dig |
| basC | reg | ~V | visa, boo | go |
| basC2; no! caution: ~w, ~x, ~Vy reg bas/homC | reg / CC | #(C)CVC | kid | bid |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #(C)CVd | kid<br>grid | bid<br>clad |
| basC2 (baking) | reg / -e | ~Ce | bake | wake |
| basC2 (freeing) | reg / e | ~ee | free | see |
| basC2 | reg / y(Cie) | ~ie | tie | lie |
| basC2 (toeing) | reg / e | ~oe | toe | shoe |
| basC2 (valuing) | reg / -e | ~ue | value | glue<br>(glu(e)ing) |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #CVg | jig | dig |
| basC2 (matches) | reg / +e | ~ch | match | catch, teach |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #CCVm | skim | swim |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #(C)CVn | sun<br>twin | win<br>spin |
| basC2 (vetoes) | reg / +e | ~Co | veto | go |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #CVp, #CCVp | cap<br>stop | kid_nap<br>wor_ship |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #(C)CVr | bar | dif_fer |
| basC2 (gassing, gasses) monosyllabic basic verbs and their prefixed verbs | reg / CC +e | #(C)CVs | gas | bus |
| basC2 monosyllabic basic verbs and their prefixed verbs | reg / CC | #(C)CVt | bat<br>flit, glut, smut | get<br>slit, spit<br>split |
| basC2 (not necessary) | reg | ~w | view | know |
| basC2 | reg / ie(Cy) | ~Cy | dry | fly |
| basC2 (not necessary) | reg | ~Vy | play | buy |

# 5.E.5 Cluster tree 1

| Level 1 | Level 2 | Level 3 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|
| L1 | | | basC | reg | ~C | look | dig |
| | L2 | | homC | reg / CC | #CVb | rob | -- |
| | L2 | | homC | reg / CC | #CCVb | crab | -- |
| | L2 | | homC | reg / CC | #CCCVb | ? | -- |
| | L2 | | homC | reg / CC | ~CVc | frolic | -- |
| | **L2** | | **homC** | **i-i-i / D** | **~ild** | **gild, build** | **--** |
| | L2 | | basC2 | reg / CC | #CVd | kid | bid |
| | L2 | | basC2 | reg / CC | #CCVd | grid | clad |
| | L2 | | homC | reg / CC | #CCCVd | ? | _-- |
| | L2 | | basC2 | reg / -e | ~Ce | bake | wake |
| | L2 | | basC2 | reg / e | ~ee | free | see |
| | L2 | | basC2 | reg / y(ie) | ~ie | tie | lie |
| | L2 | | basC2 | reg / e | ~oe | toe | shoe |
| | L2 | | basC2 | reg / -e | ~ue | value | glue |
| | **L2** | | **homC** | **i-u-u / 0** | **~ling** | **cling, fling** | **--** |
| | L2 | | basC2 | reg / CC | #CVg | jig | dig |
| | L2 | | homC | reg / CC | #CCVg | drag | -- |
| | L2 | | homC | reg / CC | #CCCVg | ? | -- |
| | L2 | | basC2 | reg / +e | ~ch | watch | teach, catch |
| | L2 | | homC | reg / +e | ~sh | fish | -- |
| | L2 | | homC | reg / CC | #CVk | ? | -- |
| | L2 | | homC | reg / CC | #CCVk | trek | -- |
| | L2 | | homC | reg / CC | #CCCVk | ? | -- |
| | L2 | | homC | reg / CC | #CVl | ex_cel, gel | -- |
| | L2 | | homC | reg / CC | #CCVl | di_stil | -- |
| | L2 | | homC | reg / CC | #CCCVl | ? | -- |
| | L2 | | homC | reg / CC | #CVm | rim | -- |
| | L2 | | basC2 | reg / CC | #CCVm | skim | swim |
| | L2 | | homC | reg / CC | #CCCVm | ? | -- |
| | L2 | | basC2 | reg / CC | #CVn | sun | win |
| | L2 | | basC2 | reg / CC | #CCVn | twin | spin |
| | L2 | | homC | reg / CC | #CCCVn | ? | -- |
| | L2 | | basC2 | reg / +e | ~Co | veto | go |
| | L2 | | homC | reg | ~oo | boo | -- |
| | L2 | | basC2 | reg / CC | #CVp | cap | kid_nap |
| | L2 | | basC2 | reg / CC | #CCVp | stop | wor_ship |
| | L2 | | homC | reg / CC | #CCCVp | strip | -- |

| Level 1 | Level 2 | Level 3 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|
| | **L2** | | **homC** | **ee-e-e / D** | **~weep** | **weep, sweep** | **--** |
| | L2 | | basC2 | reg / CC | #(C)CVr | bar | dif_fer |
| | L2 | | homC | reg / CC | #CCVr | ? | -- |
| | L2 | | homC | reg / CC | #CCCVr | ? | -- |
| | **L2** | | **homC** | **ea-o-o / n** | **~wear** | **wear, swear** | **--** |
| | L2 | | homC | reg / +e | ~ss | kiss | -- |
| | L2 | | basC2 | reg / CC +e | #CVs | gas | bus |
| | L2 | | homC | reg / CC +e | #CCVs | ? | -- |
| | L2 | | homC | reg / CC +e | #CCCVs | ? | -- |
| | L2 | | basC2 | reg / CC | #CVt | bat | get |
| | L2 | | basC2 | reg / CC | #CCVt | flit, glut, smut | slit, spit |
| | L2 | | basC2 | reg / CC | #CCCVt | ? | split |
| | L2 | | homC | reg / +e | ~x | fix | -- |
| | L2 | | basC2 | reg / ie(Cy) | ~Cy | dry | fly |
| | L2 | | homC | reg / +e | ~Cz | waltz, buzz | -- |
| | L2 | | homC | reg / CC +e | #CVz | ? | -- |
| | L2 | | homC | reg / CC +e | #CCVz | quiz | -- |
| | L2 | | homC | reg / CC +e | #CCCVz | ? | -- |
| L1 | | | basC | reg | ~V | visa, boo | go |

## Three-consonant groups at the beginning of words:

**scratch, stretch, spring**
**scl, stl, spl**
**shrink, shl**

# 5.E.5 Cluster tree 2 (not recommended)

| Level 1 | Level 2 | Level 3 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|
| L1 | | | basC | reg | ~C | look | dig |
| | L2 | | basC2 | reg / CC | #CVC | bat | get |
| | L2 | | basC2 | reg / CC | #CCVC | flit, glut, smut | slit, spit |
| | L2 | | basC2 | reg / CC | #CCCVC | strip | split |
| | ? | | homC | reg / CC | ~CVc | frolic | -- |
| | **L2** | | **homC** | **i-i-i / D** | **~ild** | **gild, build** | **--** |
| | L2 | | basC2 | reg / -e | ~Ce | bake | wake |
| | L2 | | basC2 | reg / e | ~ee | free | see |
| | L2 | | basC2 | reg / y(ie) | ~ie | tie | lie |
| | L2 | | basC2 | reg / e | ~oe | toe | shoe |
| | L2 | | basC2 | reg / -e | ~ue | value | glue |
| | **L2** | | **homC** | **i-u-u / 0** | **~ling** | **cling, fling** | **--** |
| | L2 | | basC2 | reg / +e | ~ch | watch | teach, catch |
| | L2 | | homC | reg / +e | ~sh | fish | -- |
| | L2 | | basC2 | reg / +e | ~Co | veto | go |
| | L2 | | homC | reg | ~oo | boo | -- |
| | **L2** | | **homC** | **ee-e-e / D** | **~weep** | **weep, sweep** | **--** |
| | **L2** | | **homC** | **ea-o-o / n** | **~wear** | **wear, swear** | **--** |
| | L2 | | homC | reg / +e | ~ss | kiss | -- |
| | | L3 | basC2 | reg / CC +e | #CVs | gas | bus |
| | | L3 | homC | reg / CC +e | #CCVs | ? | -- |
| | | L3 | homC | reg / CC +e | #CCCVs | ? | -- |
| | ? | | basC2 | reg | ~w | view | know |
| | ? | | homC | reg / +e | ~x | fix | -- |
| | L2 | | basC2 | reg / ie(Cy) | ~Cy | dry | fly |
| | ? | | basC2 | reg | ~Vy | play | buy |
| | L2 | | homC | reg / +e | ~Cz | waltz, buzz | -- |
| | | L3 | homC | reg / CC +e | #CVz | ? | -- |
| | | L3 | homC | reg / CC +e | #CCVz | quiz | -- |
| | | L3 | homC | reg / CC +e | #CCCVz | ? | -- |
| L1 | | | basC | reg | ~V | visa, boo | go |

# 5.E.5 Cluster tree 2 (not recommended)

**Problems:**

| | |
|---|---|
| **#(C)CVC vs. ~CVc:** | **not disjoint, no subset, same infl. types** |
| **#(C)CVC vs. ~w:** | **not disjoint, no subset, diff. infl. types** |
| **#(C)CVC vs. ~x:** | **not disjoint, no subset, diff. infl. types** |
| **#(C)CVC vs. ~Vy:** | **not disjoint, no subset, diff. infl. types** |
| | **Is *y* a consonant?** |
| **#(C)CVs** | **needed anyway (*e*-inserting)** |
| **#(C)CVz** | **needed anyway (*e*-inserting)** |

# 5.E.6 Part of the lexeme register

| Flag | Cluster | Inflection type | Analytic lex. base | Synthetic lex. base | Past tense | Past participle | Comment |
|---|---|---|---|---|---|---|---|
| s | basC | reg | ~C | ~C | ~Ced | ~Ced | e.g. look |
| s | basC | reg | ~V | ~V | ~Ved | ~Ved | e.g. visa, boo |
| s | basC2 | reg / CC | #CVg | #CVg | CVgged | CVgged | e.g. jig |
| s | homC | reg / CC | #CCVg | #CCVg | CCVgged | CCVgged | e.g. drag |
| | | reg / CC | drag | drag | dragged | dragged | |
| **s** | | **i-u-u / 0 / CC** | **dig** | **dig** | **dug** | **dug** | |
| | | reg / CC | jig | jig | jigged | jigged | |
| | | reg / CC | pig | pig | pigged | pigged | |
| s | | reg | 1hang | hang | hanged | hanged | kill with a rope |
| **v** | | **a-u-u / 0** | **2hang** | **hang** | **hung** | **hung** | suspend |
| | | reg | whang | whang | whanged | whanged | |
| | | reg | king | king | kinged | kinged | |
| **s** | **homC** | **i-u-u / 0** | **~ling** | **~ling** | **~lung** | **~lung** | **cling, fling, sling** |
| **p** | | **i-u-u / 0** | **cling** | **cling** | **clung** | **clung** | |
| **p** | | **i-u-u / 0** | **fling** | **fling** | **flung** | **flung** | |
| **p** | | **i-u-u / 0** | **sling** | **sling** | **slung** | **slung** | |
| s | | reg | 1ring | ring | ringed | ringed | provide with a ring |
| **v** | | **i-a-u / 0** | **2ring** | **ring** | **rang** | **rung** | sound |
| **s** | | **i-ou-ou / D** | **bring** | **bring** | **brought** | **brought** | |
| **s** | | **i-a-u / 0** | **α+spring** | **spring** | **sprang** | **sprung** | BE |
| **v** | | **i-a-u / 0** | **β+spring** | **spring** | **sprung** | **sprung** | AE |
| **s** | | **i-u-u / 0** | **string** | **string** | **strung** | **strung** | |
| **s** | | **i-u-u / 0** | **wring** | **wring** | **wrung** | **wrung** | |
| **s** | | **i-a-u / 0** | **sing** | **sing** | **sang** | **sung** | |
| | | reg | ting | ting | tinged | tinged | |
| **s** | | **i-u-u / 0** | **sting** | **sting** | **stung** | **stung** | |
| | | reg | wing | wing | winged | winged | |
| **s** | | **i-u-u / 0** | **swing** | **swing** | **swung** | **swung** | |
| | | reg | catalog | catalog | cataloged | cataloged | |
| s | | reg / CC | hum_bug | humbug | humbugged | humbugged | |
| s | basC2 | reg / +e | ~ch | ~ch | ~ched | ~ched | e.g. reach |
| | | reg / +e | reach | reach | reached | reached | |
| **s** | | **ea-au-au / D / +e** | **teach** | **teach** | **taught** | **taught** | |
| | | reg / +e | screech | screech | screeched | screeched | |
| s | | reg / +e | α+be_seech | beseech | beseeched | beseeched | |
| **v** | | **ee-ou-ou / D** | **β+be_seech** | **beseech** | **besought** | **besought** | |
| | | reg / +e | batch | batch | batched | batched | |
| **s** | | **a-au-au / D / +e** | **catch** | **catch** | **caught** | **caught** | |
| | | reg / +e | scratch | scratch | scratched | scratched | |
| | | reg / +e | watch | watch | watched | watched | |

**(cf. Holl / Maroldo / Urban 2007, 114-115)**

# 5.E.7 Result

**There is a regular (weak, productive) and
an irregular (weak and strong, not productive) conjugation class
(e.g. *learn, learnt, learnt; dig, dug, dug*)
with a very restricted categorisation
via reverse sorting of the present infinitives.**

**The regular conjugation type
contains productive subtypes with regular adaptations
due to phonotactic, phonetic and orthographic particularities:
consonant doubling (*digging*), *e*-insertion (*wishes*),
*e*-deletion (*aging*), *ie-y* change (*dying*) and *y-ie* change (*studies*).**

**There are only few starting points
for analogical reasoning based upon reverse similarity.
Only one twentieth (9/175) of the irregular verbs can be found
in homogeneous clusters of reversely similar verbs
with an average size of 2 (9/4).**

# 5.S The situation of the Swedish verb

## 5.S.1 Principal parts

**There are 3 to 4 principal parts depending on the regularity of present tense.**
**– infinitive**
**– (present tense)**
**– past tense**
**– supine**

**There are only a couple of forms to be derived from the principal parts:**

**imperative and present tense from the infinitive,**
**passive voice forms from the corresponding active voice forms**

**The derivation rules run approximately:**

**imperative:** **inf. - *V* + *a* / - / *V***
**present tense of active voice:** **inf. - *V* + *ar* / *er* / *Vr***
**present tense of passive voice:** **p.t. - *Vr* / *er* + *Vs* / *s***
**other tenses of passive voice:** **form + *s***

**The derivation rules depend on the conjugation type**
**which is determined by the supine and,**
**in case of its irregularity, by the present tense of active voice.**

# 5.S.2 Homogeneous clusters

| Infl. type | Lex. base | Examples |
|---|---|---|
| **4** | **~lida** | **lida, glida** |
| **4** | **~rida** | **rida, skrida, sprida (äv. 2), strida, vrida** |
| **4** | **~juda** | **bjuda, ljuda (äv. 1), sjuda** |
| 2 | ~yda | förhyda, lyda (äv. 24), pryda, tyda |
| 2 | ~råda | råda, tråda |
| 2 | ~länga | länga, blänga, flänga, klänga, slänga |
| 2 | ~ränga | kränga, spränga, tränga, stränga (äv. 1), vränga |
| 2 | ~röja | röja, dröja |
| 2 | ~rycka | rycka, trycka |
| 2 | ~räcka | räcka, bräcka, förskräcka, spräcka, träcka (äv. 1), sträcka |
| 2 | ~täcka | täcka, stäcka |
| 2 | ~leka | leka, bleka |
| **4** | **~vika** | **vika (äv. 42), svika** |
| 2 | ~ränka | dränka, kränka, skränka |
| 2 | ~tänka | tänka, stänka |
| 2 | ~ärka | märka, stärka, värka |
| 2 | ~läka | läka, fläka |
| 2 | ~räka | bräka, kräkas, vräka |
| 2 | ~röka | röka, kröka |
| 2 | ~rälla | drälla, skrälla |
| 2 | ~rämma | drämma, skrämma |
| 2 | ~änna | känna, nännas, spänna, ränna, bränna |
| 2 | ~älpa | hjälpa, stjälpa |
| 2 | ~näppa | knäppa, snäppa |
| 2 | ~ärpa | skärpa, snärpa, värpa |
| 2 | ~öpa | döpa, köpa, löpa snöpa, gröpa (äv. 1), stöpa |
| 2 | ~väsa | väsa, kväsa |
| **4** | **~juta** | **gjuta, skjuta, ljuta, njuta, tjuta** |
| **4** | **~nyta** | **knyta, snyta** |
| **4** | **~ryta** | **ryta, bryta, skryta, tryta** |
| 2 | ~älva | skälva (äv. 24), välva |
| 3 | ~o | bo, sko, varsko, glo, gno, sno, ro, gro, tro |
| 3 | ~y | sky, lyss, fly, gny, spy, bry, gry, sy, ty |
| 3 | ~rå | rå, brås, förebrå, trå |
| 3 | ~ä | klä, spä, trä |

# 5.S.2 Homogeneous clusters

## 2nd conjugation

***-yda****: förhyda (förhydde förhytt), lyda (auch 4.), pryda, tyda*

***-råda****: råda (rådde rått), tråda*

***-länga****: länga (längde längt), blänga, flänga, klänga, slänga*

***-ränga****: kränga (krängde krängt), spränga, tränga, stränga* (auch 1. mit Bedeutungsunterschied)*, vränga*

***-änga*** (weitgehend homogen)*:* zu *-länga* und *-ränga* kommen hinzu *dänga, hänga, mänga, stänga, svänga;* Ausnahmen *gänga, mannekänga* (beide 1.)

***-röja****: röja (röjde röjt), dröja*

***-öja*** (fast homogen)*:* zu *-röja* kommen hinzu *böja, höja, flöja, plöja, nöja, töja;* Ausnahme *slöja* (1.)

***-rycka****: rycka (ryckte ryckt), trycka*

***-räcka****: räcka (räckte räckt), bräcka, förskräcka, spräcka, träcka* (auch 1. mit Bedeutungsunterschied)*, sträcka*

***-täcka****: täcka (täckte täckt), stäcka*

***-leka****: leka (lekte lekt), bleka*

***-ränka****: dränka (dränkte dränkt), kränka, skränka*

***-tänka****: tänka (tänkte tänkt), stänka*

***-änka*** (weitgehend homogen)*:* zu *-ränka* und *-tänka* kommen hinzu *dänka, skänka, blänka, sänka;* Ausnahmen *bänka, länka* (beide 1.)

***-ärka****: märka (märkte märkt), stärka, värka*

***-läka**: läka (läkte läkt), fläka*

***-räka**: bräka (bräkte bräkt), kräkas* (Deponens)*, vräka*

***-röka**: röka (rökte rökt), kröka*

***-rälla**: drälla (drällde drällt), skrälla*

***-rämma**: drämma (drämde drämt), skrämma*

***-änna**: känna (kände känt), nännas* (Deponens)*, spänna, ränna, bränna*

***-älpa**: hjälpa (hjälpte hjälpt), stjälpa*

***-näppa**: knäppa (knäppte knäppt), snäppa*

***-äppa*** (fast homogen)*:* zu *-näppa* kommen hinzu *släppa, skräppa, täppa;* Ausnahme *kräppa (1.)*

***-ärpa**: skärpa (skärpte skärpt), snärpa, värpa*

***-öpa**: döpa (döpte döpt), köpa, löpa, snöpa, gröpa* (auch 1. mit Bedeutungsunterschied)*, ströpa*

***-väsa**: väsa (väste väst), kväsa*

***-älva**: skälva (skälvde skälvt), välva*

## 4th conjugation

***-lida**: lida (led lidit), glida*

***-rida**: rida (red ridit), skrida, sprida* (auch 2.)*, strida, vrida*

***-juda**: bjuda (bjöd bjudit), sjuda, ljuda* (1|4)

***-vika**: vika (vek vikit), svika*

***-rypa*** (fast homogen)*: drypa (dröp drupit), krypa, strypa* (2|24))

***-juta**: gjuta (göt gjutit), skjuta, ljuta, njuta, tjuta*

***-nyta**: knyta (knöt knutit), snyta*

***-ryta**: ryta (röt rutit), bryta, skryta, tryta*

***-yta*** (fast homogen)*:* zu *-ryta* kommen hinzu *flyta, knyta, snyta;* Ausnahme *byta (2.)*

# 5.S.3 Inhomogeneous clusters

**1st and 2nd conjugation: *-eda***
***leda (ledade, ledat)*** ‘böja’*, skeda, reda* ‘rederi’*, freda*
***leda (ledde, lett)*** ‘föra’*, reda, breda, sveda*

**1st and 4th conjugation: *-inda***
***vinda (vindade, vindat),***
***binda (band, bundit),***

**1st, 2nd and 4th conjugation: *-ida***
***snida (snidade, snidat)****, guida, svida* ‘klä’
***smida (smidde, smitt),*** *idas, sprida*
***lida (led, lidit),*** *glida, nida, rida, skrida, sprida, (be)strida, vrida, kvida, svida* ‘göra ont’

**2nd and 4th conjugation**
***byta (bytte, bytt)***
***flyta (flöt, flutit),*** *knyta, snyta, ryta, bryta, skryta, tryta*

**3rd and 4th conjugation**
***må (mådde, mått)*** *‘känna sig’, flå, klå, försmå, nå, spå, rå, brås, förebrå, trå, så*
***må (måtte, mått)*** *‘hjälpverb’*
***stå (stod, stått)***
***få (fick, fått),*** *gå*
***slå (slog, slagit)***

**etc.**

# 5.S.4 Basic clusters

**Only two:**

***~a (~ar, ~ade, ~at)***
***~as (~as, ~ades, ~ats)***

**The 1st conjugation and its deponentia.**
**The 1st conjugation is the only productive conjugation type.**

# 5.S.5 Cluster tree

| Le-vel 1 | Le-vel 2 | Le-vel 3 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|
| L1 | | | basC | 1 | ~a | börja | äta |
| | **L2** | | **homC** | **4** | **~lida** | **lida** | **--** |
| | **L2** | | **homC** | **4** | **~rida** | **rida** | **--** |
| | **L2** | | **homC** | **4** | **~juda** | **bjuda** | **--** |
| | L2 | | homC | 2 | ~yda | tyda | -- |
| | L2 | | homC | 2 | ~råda | råda | -- |
| | L2 | | homC | 2 | ~länga | länga | -- |
| | L2 | | homC | 2 | ~ränga | ränga | -- |
| | L2 | | homC | 2 | ~röja | röja | -- |
| | L2 | | homC | 2 | ~rycka | trycka | -- |
| | L2 | | homC | 2 | ~räcka | räcka | -- |
| | L2 | | homC | 2 | ~täcka | täcka | -- |
| | L2 | | homC | 2 | ~leka | leka | -- |
| | **L2** | | **homC** | **4** | **~vika** | **vika** | **--** |
| | L2 | | homC | 2 | ~ränka | dränka | -- |
| | L2 | | homC | 2 | ~tänka | tänka | -- |
| | L2 | | homC | 2 | ~ärka | märka | -- |
| | L2 | | homC | 2 | ~läka | läka | -- |
| | L2 | | homC | 2 | ~räka | bräka | -- |
| | L2 | | homC | 2 | ~röka | röka | -- |
| | L2 | | homC | 2 | ~rälla | drälla | -- |
| | L2 | | homC | 2 | ~rämma | drämma | -- |
| | L2 | | homC | 2 | ~änna | känna | -- |
| | L2 | | homC | 2 | ~älpa | hjälpa | -- |
| | L2 | | homC | 2 | ~näppa | knäppa | -- |
| | L2 | | homC | 2 | ~ärpa | skärpa | -- |
| | L2 | | homC | 2 | ~öpa | döpa | -- |
| | L2 | | homC | 2 | ~väsa | väsa | -- |
| | **L2** | | **homC** | **4** | **~juta** | **gjuta** | **--** |
| | **L2** | | **homC** | **4** | **~nyta** | **knyta** | **--** |
| | **L2** | | **homC** | **4** | **~ryta** | **ryta** | **--** |
| | L2 | | homC | 2 | ~älva | välva | -- |
| | L2 | | homC | 3 | ~o | bo | -- |
| | L2 | | homC | 3 | ~y | bry | -- |
| | L2 | | homC | 3 | ~rå | rå | -- |
| | L2 | | homC | 3 | ~ä | klä | -- |
| L1 | | | basC | 1 | ~as | kallas | finnas |

# 5.S.6 Part of the lexeme register

| Flag | Cluster | Infl. type | Analyt. lex. base | Synthet. lex.base | Pres. tense | Past tense | Supine | Comment |
|---|---|---|---|---|---|---|---|---|
| s | basC | 1 | ~a | ~a | | ~ade | ~at | |
| | | 1 | skeda | skeda | | skedade | skedat | |
| s | | 1 | 1leda | leda | | ledade | ledat | röra sig i en led, böja |
| v | | 2 | 2leda | leda | | ledde | lett | föra, vara främst |
| v | | 2 | (s) *2leda | ledas | | leddes | letts | känna leda |
| s | | 1 | 1reda | reda | | redade | redat | idka rederigörelse |
| v | | 2 | 2reda | reda | | redde | rett | göra i ordning etc. |
| s | | 2 | breda | breda | | bredde | brett | |
| | | 1 | freda | freda | | fredade | fredat | |
| s | | 2 | sveda | sveda | | svedde | svett | |
| s | | 2 | (s) *ida | idas | | iddes | itts | |
| | | 1 | skida | skida | | skidade | skidat | |
| **s** | **homC** | **4** | **~lida** | **~lida** | | **~led** | **~lidit** | |
| s | | 2 | smida | smida | | smidde | smitt | |
| **s** | | **4** | **gnida** | **gnida** | | **gned** | **gnidit** | |
| | | 1 | snida | snida | | snidade | snidat | |
| **s** | **homC** | **4** | **~rida** | **~rida** | | **~red** | **~ridit** | |
| **s** | | **4** | **α+sprida** | **sprida** | | **spred** | **spridit** | |
| v | | 2 | β+sprida | sprida | | (spridde) | spritt | |
| **s** | | **4** | **1strida** | **strida** | | **stred** | **stridit** | **kämpa, tvista** |
| **s** | | **4** | **α+be_2strida** | **bestrida** | | **bestred** | **bestridit** | **tillbakavisa, förneka** |
| **v** | | **42** | **β+be_2strida** | **bestrida** | | **bestred** | **bestritt** | **tillbakavisa, förneka** |
| | | 1 | guida | guida | | guidade | guidat | |
| **s** | | **4** | **kvida** | **kvida** | | **kved** | **kvidit** | |
| s | | 1 | 1svida | svida | | svidade | svidat | klä |
| **s** | | **4** | **2svida** | **svida** | | **sved** | **svidit** | **göra ont** |
| **s** | | **4** | **binda** | **binda** | | **band** | **bundit** | |
| | | 1 | vinda | vinda | | vindade | vindat | |
| | | 1 | ända | ända | | ändade | ändat | |
| | | 1 | full_ända | fullända | | fulländade | fulländat | |
| s | | 2 | bända | bända | | bände | bänt | |
| s | | 2 | hända | hända | | hände | hänt | |
| | | 1 | skända | skända | | skändade | skändat | |
| s | | 2 | lända | lända | | lände | länt | |
| | | 1 | blända | blända | | bländade | bländat | |
| s | | 2 | sända | sända | | sände | sänt | |
| s | | 2 | tända | tända | | tände | tänt | |
| s | | 2 | vända | vända | | vände | vänt | |
| s | | 1 | (1varda | varda | | vardade | vardat | [inte i SAOL] |
| s | | 1 | ant_1varda | antvarda | | antvardade | antvardat | |
| s | | 2 | 2varda | varda | varder | vart | (vorden) | |

# 5.S.7 Result

**There is a regular weak (1, productive),**
**another regular weak (3),**
**an irregular weak (2),**
**a mixed and an irregular strong conjugation class (4)**
**with a very restricted categorisation**
**via reverse sorting of the present infinitives.**

**A clearer categorization can only be achieved**
**with the help of a second feature, the ending of the supine**
**(a neutrum-like form of the past participle**
**used for the synthetic tenses of the past).**

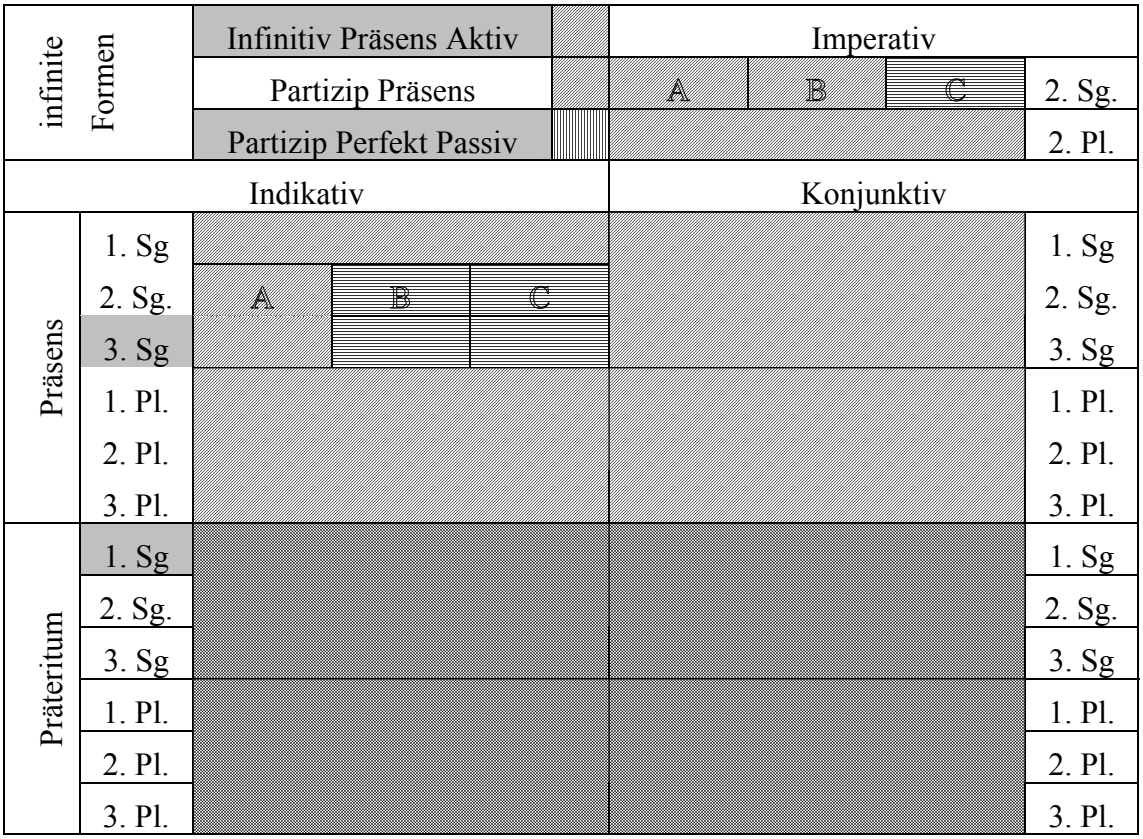
**There are grey zones due to verbs**
**which simultaneously have features of different classes.**

**There are numerous starting points**
**for analogical reasoning based upon reverse similarity.**

**About one fifth (92/460) of the irregular verbs can be found**
**in homogeneous clusters of reversely similar verbs**
**with an average size of 3 (92/30).**

# 5.D The situation of the German verb

## 5.D.1 Principal parts




**A**: Indikativ Präsens 2. / 3. Sg. und Imperativ 2. Sg. leiten sich vom Infinitiv ab. Indikativ Präsens 3. Sg. ist keine Schlüsselform.

*gehen – du gehst – er geht – geh!*

**B**: Indikativ Präsens 2. Sg. leitet sich vom Indikativ Präsens 3. Sg. (Schlüsselform) ab, der Imperativ 2. Sg. aber vom Infinitiv.

*lassen – du lässt – er lässt – lass!*

**C**: Indikativ Präsens 2. Sg. und Imperativ 2. Sg. leiten sich vom Indikativ Präsens 3. Sg. ab, der als Schlüsselform genannt ist.

*treffen – du triffst – er trifft – triff!*

**Stem distribution of the German verb**
**(Holl / Behrschmidt / Kühn 2004, 103-104)**

# 5.D.2 Homogeneous clusters

| Inflection type | Lex. base | Examples |
|---|---|---|
| **ei-ie-ie** | **~reiben** | **reiben, schreiben, treiben** |
| **i-a-u / +e(dt)** | **~winden** | **winden, schwinden** |
| **i-a-u** | **~lingen** | **schlingen, klingen, ge- / miss_lingen** |
| **i-a-u** | **~wingen** | **schwingen, zwingen** |
| **ie-o-o** | **~riechen** | **riechen, kriechen** |
| **ei-i-i** | **~leichen** | **bleichen, gleichen, schleichen** |
| reg / +e(mn) | ~dmen | widmen |
| **i-o-o** | **~limmen** | **glimmen, klimmen** |
| reg / +e(mn) | ~tmen | atmen |
| reg / +e(mn) | ~bnen | ebnen |
| reg / +e(mn) | ~dnen | ordnen |
| reg / +e(mn) | ~fnen | öffnen |
| reg / +e(mn) | ~gnen | eignen |
| reg / +e(mn) | ~chnen | rechnen |
| reg / +e(mn) | ~cknen | trocknen |
| reg / +e(mn) | ~mnen | ver_vollkommnen |
| reg / +e(mn) | ~pnen | wappnen |
| **ie-o-o / -s / +e(sz) / +CC** | **~ließen** | **fließen, schließen** |
| **ie-o-o / -s / +e(sz) / +CC** | **~rießen** | **sprießen, ver_drießen** |
| reg / -s(sz) | ~xen | faxen |
| reg / -e(lr) | ~eln | jubeln |
| reg / -e(lr) | ~ern | zaubern |

# 5.D.2 Homogeneous clusters

***~reiben**: reiben (rieb gerieben), schreiben, treiben*

***~winden**: winden (windet wand gewunden), schwinden*

***~lingen**: schlingen (schlang geschlungen), klingen, ge-/miss_lingen*

***~wingen**: schwingen (schwang geschwungen), zwingen*

***~riechen**: riechen (roch gerochen), kriechen*

***~leichen**: bleichen (blich geblichen)* (auch regelmäßig), *gleichen, schleichen*

***~limmen**: glimmen (glomm, geglommen)* (auch regelmäßig), *klimmen* (auch regelmäßig)

***~ließen**: fließen (floss geflossen), schließen*

***~rießen**: sprießen (spross gesprossen)* (auch regelmäßig), *verdrießen*

# 5.D.3 Inhomogeneous clusters 1

**Necessarily contain irregular verbs**

**Selection of clusters with 3-4 inflection types**

***graben (grub, gegraben)***
***haben (hatte, gehabt)***
***laben (labte, gelabt), schaben, traben***

***schneiden (schnitt, geschnitten), leiden***
***meiden (mied, gemieden), scheiden***
***kleiden (kleidete, gekleidet)***

***liegen (lag gelegen)***
***biegen (bog gebogen), fliegen, wiegen***
***siegen (siegte gesiegt), kriegen, schmiegen***

***gehen (ging, gegangen)***
***stehen (stand, gestanden)***
***sehen (sah, gesehen), geschehen***
***flehen (flehte, gefleht), drehen***

***schwimmen (schwamm, geschwommen)***
***glimmen (glomm - glimmte, geglommen - geglimmt), klimmen***
***stimmen (stimmte, gestimmt), grimmen, trimmen***

***heißen (hieß, geheißen)***
***beißen (biss, gebissen), scheißen, schleißen, schmeißen, reißen***
***spleißen (spliss - spleißte, gesplissen, gespleißt)***
***weißen (weißte, geweißt), gleißen, kreißen, schweißen***

**etc.**

# 5.D.3 Inhomogeneous clusters 2

## Clusters with 2 inflection types only are considerably more frequent

***~eiben*** (fast homogen): zu *~reiben* kommt hinzu *bleiben*; Ausnahme *einver-/ent_leiben* (regelmäßig)

***~ingen*** (weitgehend homogen): zu *~lingen* und *~wingen* kommen hinzu *dingen* (auch regelmäßig), *ringen, dringen, springen, wringen, singen*; Ausnahmen *bringen (brachte gebracht), be-/um_ringen* (regelmäßig)

***~ießen*** (fast homogen): zu *~ließen* und *~rießen* kommen hinzu *gießen, schießen, genießen*; Ausnahme *spießen* (regelmäßig)

**etc.**

# 5.D.4 Basic clusters

| Cluster information | Basic cluster: morphological property (infl. type) | Basic cluster: alphabetic property (lexical base) | Examples | Exceptions |
|---|---|---|---|---|
| basC | reg | ~en | loben | gehen |
| basC2 (er badet) | reg / +e(dt) | ~den | baden | winden |
| basC2; no! caution | reg / +e(mn) | ~Cmen | atmen | formen, kommen |
| basC2; no! caution | reg / +e(mn) | ~Cnen | ebnen | lernen, kennen |
| basC2 (hat studiert) | reg / -ge(ieren) | ~ieren | studieren | stieren |
| basC2 (du grast) | reg / -s(sz) | ~sen | grasen, passen | blasen, essen |
| basC2 (du spaßt) | reg / -s(sz) | ~ßen | spaßen | fließen |
| basC2 (er watet) | reg / +e(dt) | ~ten | waten | raten |
| basC2 (du pflanzt) | reg / -s(sz) | ~zen | pflanzen | sitzen |

**basC**
***~en (~t, ~te, ge~t)***

**basC2**
***~den (~det, ~dete, ge~det): baden***
***Cmen (~Cmet, ~Cmete, ge~Cmet): atmen***
***Cnen (~Cnet, ~Cnete, ge~Cnet): öffnen***
***~ieren (~iert, ~ierte, ~iert): studieren***
***~sen***
***~ßen***
***~ten (~tet, ~tete, ~tet)***
***~zen***

# 5.D.5 Cluster tree 1

| Level 1 | Level 2 | Level 3 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|
| L1 | | | basC | reg | ~en | loben | gehen |
| | **L2** | | **homC** | **ei-ie-ie** | **~reiben** | **reiben** | **--** |
| | L2 | | basC2 | reg / +e(dt) | ~den | baden | winden |
| | | **L3** | **homC** | **i-a-u / +e(dt)** | **~winden** | **winden** | **--** |
| | **L2** | | **homC** | **i-a-u** | **~lingen** | **gelingen** | **--** |
| | **L2** | | **homC** | **i-a-u** | **~wingen** | **schwingen** | **--** |
| | **L2** | | **homC** | **ie-o-o** | **~riechen** | **riechen** | **--** |
| | **L2** | | **homC** | **ei-i-i** | **~leichen** | **gleichen** | **--** |
| | L2 | | homC | reg / +e(mn) | ~dmen | widmen | -- |
| | **L2** | | **homC** | **i-o-o** | **~limmen** | **glimmen** | **--** |
| | L2 | | homC | reg / +e(mn) | ~tmen | atmen | -- |
| | L2 | | homC | reg / +e(mn) | ~bnen | ebnen | -- |
| | L2 | | homC | reg / +e(mn) | ~dnen | ordnen | -- |
| | L2 | | homC | reg / +e(mn) | ~fnen | öffnen | -- |
| | L2 | | homC | reg / +e(mn) | ~gnen | eignen | -- |
| | L2 | | homC | reg / +e(mn) | ~chnen | rechnen | -- |
| | L2 | | homC | reg / +e(mn) | ~cknen | trocknen | -- |
| | L2 | | homC | reg / +e(mn) | ~mnen | ver_vollkommnen | -- |
| | L2 | | homC | reg / +e(mn) | ~pnen | wappnen | -- |
| | L2 | | basC2 | reg / -ge(ieren) | ~ieren | studieren | stieren |
| | L2 | | basC2 | reg / -s(sz) | ~sen | grasen | blasen |
| | L2 | | basC2 | reg / -s(sz) | ~ßen | spaßen | fließen |
| | | **L3** | **homC** | **ie-o-o / -s/+e(sz)/+CC** | **~ließen** | **fließen** | **--** |
| | | **L3** | **homC** | **ie-o-o / -s/+e(sz)/+CC** | **~rießen** | **sprießen** | **--** |
| | L2 | | basC2 | reg / +e(dt) | ~ten | waten | raten |
| | L2 | | homC | reg / -s(sz) | ~xen | faxen | -- |
| | L2 | | basC2 | reg / -s(sz) | ~zen | pflanzen | sitzen |
| L1 | | | homC | reg / -e(lr) | ~eCn | handeln, zaubern | -- |

# 5.D.5 Cluster tree 2 (not recommended)

**basC2 ~Cmen (reg / +e(mn)) would require**
**basC3 ~hmen (reg), homC ~lmen (reg),**
**basC3 ~mmen (reg), homC ~rmen (reg)**

**basC2 ~Cnen (reg / +e(mn)) would require**
**homC ~Vhnen (reg), basC3 ~nnen (reg), homC ~rnen (reg)**

| Level 1 | Level 2 | Level 3 | Level 4 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|---|
| | L2 | | | basC2 | reg / +e(mn) | ~Cmen | atmen | kommen |
| | | L3 | | basC3 | reg | ~hmen | nach_ahmen | nehmen |
| | | L3 | | homC | reg | ~lmen | qualmen | -- |
| | | L3 | | basC3 | reg | ~mmen | stammen | glimmen |
| | | | **L4** | **homC** | **i-o-o** | **~limmen** | **glimmen** | **--** |
| | | L3 | | homC | reg | ~rmen | formen | -- |
| | L2 | | | basC2 | reg / +e(mn) | ~Cnen | ebnen | kennen |
| | | L3 | | homC | reg | ~Vhnen | dehnen | -- |
| | | L3 | | basC3 | reg | ~nnen | sonnen | nennen |
| | | L3 | | homC | reg | ~rnen | lernen | -- |

## Part of the cluster tree

# 5.D.6 Part of the lexeme register

| Flag | Cluster | Infl. type | Analytic lex. base | Synthet. lex. base | Present 3rd sg | Past tense 1st sg | Past Participle | Comment |
|---|---|---|---|---|---|---|---|---|
| s | basC | reg | ~en | ~en | | ~te | ge~t | e.g. loben |
| | | reg | säen | säen | | säte | gesät | |
| **s** | | **a-a-a/Aux** | **haben** | **haben** | **hat** | **hatte** | **gehabt** | |
| **s** | | **a-a-a/Aux** | **inne_haben** | **innehaben** | **hat inne** | **hatte inne** | **innegehabt** | |
| p | | reg | schaben | schaben | | schabte | geschabt | |
| s | | reg | handhaben | handhaben | | hand-habte | gehandhabt | noun Handhabe |
| **s** | | **a-ä-u-a** | **graben** | **graben** | **gräbt** | **grub** | **gegraben** | |
| p | | reg | traben | traben | | trabte | getrabt | |
| **s** | | **e-i-a-e** | **geben** | **geben** | **gibt** | **gab** | **gegeben** | **imper. gib** |
| **s** | | **e-o-o** | **heben** | **heben** | | **hob** | **gehoben** | |
| | | **e-o-o** | **auf_heben** | **aufheben** | **hebt auf** | **hob auf** | **aufgehoben** | |
| **s** | | **ie-o-o** | **schieben** | **schieben** | | **schob** | **geschoben** | |
| p | | reg | lieben | lieben | | liebte | geliebt | |
| s | | reg | α+stieben | stieben | | stiebte | gestiebt | |
| **v** | | **ie-o-o** | **β+stieben** | **stieben** | | **stob** | **gestoben** | |
| | | reg | leben | leben | | lebte | gelebt | |
| s | | reg | α+weben | weben | | webte | gewebt | |
| **v** | | **e-o-o** | **β+weben** | **weben** | | **wob** | **gewoben** | |
| p | | reg | schweben | schweben | | schwebte | geschwebt | |
| p | | reg | einver_leiben | einverleiben | | verleibte ein | einverleibt | |
| s | | ei-ie-ie | bleiben | bleiben | | blieb | geblieben | |
| **s** | **homC** | **ei-ie-ie** | **~reiben** | **~reiben** | | **~rieb** | **ge~rieben** | |
| **p** | | **ei-ie-ie** | **reiben** | **reiben** | | **rieb** | **gerieben** | |
| **p** | | **ei-ie-ie** | **schreiben** | **schreiben** | | **schrieb** | **geschrieben** | |
| **p** | | **ei-ie-ie** | **treiben** | **treiben** | | **trieb** | **getrieben** | |
| p | | reg | erben | erben | | erbte | geerbt | |
| **s** | | **e-i-a-o** | **ver_derben** | **verderben** | **verdirbt** | **verdarb** | **verdorben** | **imper. verdirb** |
| p | | reg | gerben | gerben | | gerbte | gegerbt | |
| p | | reg | kerben | kerben | | kerbte | gekerbt | |
| **s** | | **e-i-a-o** | **sterben** | **sterben** | **stirbt** | **starb** | **gestorben** | **imper. stirb, subj. past stürbe** |
| **s** | | **e-i-a-o** | **werben** | **werben** | **wirbt** | **warb** | **geworben** | **imper. wirb** |
| s | | reg | α+schnauben | schnauben | | schnaubte | geschnaubt | current use |
| **v** | | **au-o-o** | **β+schnauben** | **schnauben** | | **schnob** | **geschnoben** | **obsolete** |
| p | | reg | rauben | rauben | | raubte | geraubt | |
| s | basC2 | reg/+e(dt) | ~den | ~den | ~det | ~dete | ge~det | e.g. baden |
| p | | reg/+e(dt) | baden | baden | badet | badete | gebadet | |
| **s** | | **a-ä-u-a/+e(dt)** | **laden** | **laden** | **lädt** | **lud** | **geladen** | |

# 5.D.7 Result

**Es gibt es eine regelmäßige (schwache, produktive) und eine unregelmäßige (starke, nicht produktive) Konjugationsklasse ohne durchgängige Abgrenzungsmöglichkeit über rückläufige Sortierung.**

**Die reguläre Konjugationsklasse**
**enthält produktive Subtypen mit regelhaften Anpassungen aufgrund von phonotaktischen, phonetischen und orthographischen Besonderheiten:**

***e*-Einschübe bei Verben auf *-den, -ten, -Cmen* (*C* ungleich *h, l, m, r*; alle homogen), *-Cnen* (nur *-fnen, -gnen, -chnen* homogen)**

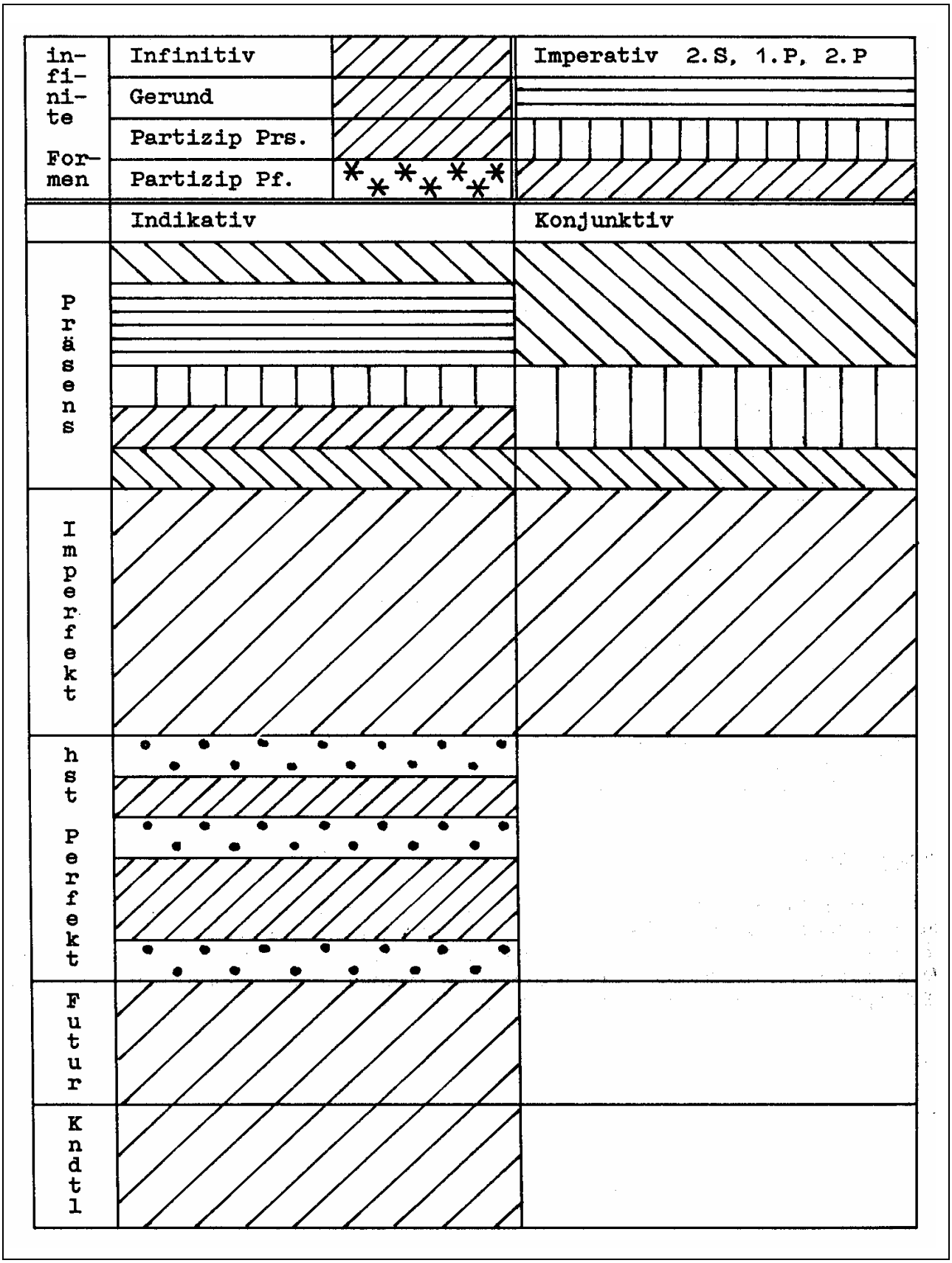
***e*-Elision im Ind.Prs.1.Sg. bei Verben auf *-eln* und *-ern* (beide homogen)**

***s*-Elision im Ind.Prs.2.Sg. bei Verben auf *-sen, -ssen, -ßen, -xen* (homogen), *-zen*.**

**Ansätze für ausgangsbasierte Analogieschlüsse sind vorhanden, aber nicht zahlreich.**
**Etwa ein Neuntel (21/190) der unregelmäßigen Verben**
**liegt in homogenen Gruppen ausgangsgleicher**
**mit der Durchschnittsgröße 2 ½ (21/9).**

# 5.I The situation of the Italian verb

## 5.I.1 Principal parts 1



# 5.I.1 Principal parts 2

**Principal parts:**
**Inf.Prs.,**
**Ind.Prs.1.Sg.,**
**Ind.Prs.2.Sg.,**
**Ind.Prs.1.Pl.,**
**hist.Pf.1.Sg.,**
**Past Ptcp.,**
**Fut.1.Sg.**

**The derivation rules depend on the conjugation class which is determined by the infinitive ending.**

**(Holl 1988, 232, 234-235)**

# 5.I.1 Principal parts 3

**Ind.Prs.**

| | Inf.Prs.2 auf -Vre<br>V = a | auf -Vre<br>V = e i |
|---|---|---|
| 1.S | Schlüsselform o | o |
| 2.S | Schlüsselform i | i |
| 3.S | Ind.Prs.2.S - i + a | - i + e |
| 1.P | Schlüsselform iamo | iamo |
| 2.P | Inf.Prs.2 - Vre + Vte | - Vre + Vte |
| 3.P | Ind.Prs.1.S - o + ano | - o + ono |

**Ipv.**

| | | |
|---|---|---|
| 2.S | Ind.Prs.2.S - i + a | i |

| | | |
|---|---|---|
| 1.P | Ind.Prs.1.P | iamo |
| 2.P | Ind.Prs.2.P | Vte |

Inf.Prs.1 Inf.Prs.2

**Ind.Ipf.; V = a e i**

| | | | | | |
|---|---|---|---|---|---|
| 1.S | Inf.Prs.2 | - | re | + | vo |
| 2.S | Ind.Ipf.1.S | - | o | + | i |
| 3.S | Ind.Ipf.1.S | - | o | + | a |
| 1.P | Ind.Ipf.1.S | - | o | + | amo |
| 2.P | Ind.Ipf.1.S | - | o | + | ate |
| 3.P | Ind.Ipf.1.S | - | Vvo | + | Vvano |

**hist.Perf.**

| | hist.Pf.1.S auf -Vi | h.Pf.1.S auf -Ci |
|---|---|---|
| 1.S | Schlüsselform Vi | Ci |
| 2.S | Inf.Prs.2 - Vre + Vsti | - Vre + Vsti |
| 3.S | h.Pf.1.S - Vi + X | - Ci + Ce |
| 1.P | h.Pf.2.S - Vsti + Vmmo | - Vsti + Vmmo |
| 2.P | h.Pf.2.S - Vsti + Vste | - Vsti + Vste |
| 3.P | h.Pf.1.S - Vi + Vrono | - Ci + Cero |

X = ò V = a
X = V‘ V = e i u

**Futur**

| | | | | |
|---|---|---|---|---|
| 1.S | Inf.Prs.1 | | + | ò |
| 2.S | Fut.1.S | - ò | + | ai |
| 3.S | Fut.1.S | - ò | + | à |
| 1.P | Fut.1.S | - ò | + | emo |
| 2.P | Fut.1.S | - ò | + | ete |
| 3.P | Fut.1.S | - ò | + | anno |

# 5.I.1 Principal parts 4

**Konj.Prs.**

Inf.Prs.2 auf -are : W = i
Inf.Prs.2 auf -ere, -ire : W = a

| | | | | |
|---|---|---|---|---|
| 1.S | Ind.Prs. 1.S | - o | + | W |
| 2.S | Konj.Prs.1.S | | | W |
| 3.S | Konj.Prs.1.S | | | W |
| 1.P | Ind.Prs. 1.P | | | iamo |
| 2.P | Ind.Prs. 1.P | - iamo | + | iate |
| 3.P | Konj.Prs.1.S | - W | + | Wno |

**Part.Prs.**

| | | | |
|---|---|---|---|
| Inf.Prs.2 | - Vre | + Vnte | V = a |
| Inf.Prs.2 | - Vre | + ente | V = e i |

**Gerund**

| | | | |
|---|---|---|---|
| Inf.Prs.2 | - Vre | + Vndo | V = a |
| Inf.Prs.2 | - Vre | + endo | V = e i |

**Konj.Ipf.;** V = a e i o

| | | | | |
|---|---|---|---|---|
| 1.S | hist.Pf.2.S | - ti | + | si |
| 2.S | hist.Pf.2.S | - ti | + | si |
| 3.S | hist.Pf.2.S | - ti | + | se |
| 1.P | hist.Pf.2.S | - Vsti | + | Vssimo |
| 2.P | hist.Pf.2.S | - ti | + | te |
| 3.P | hist.Pf.2.S | - Vsti | + | Vssero |

**Konditional**

| | | | | |
|---|---|---|---|---|
| 1.S | Fut.1.S | - ò | + | ei |
| 2.S | Fut.1.S | - ò | + | esti |
| 3.S | Fut.1.S | - ò | + | ebbe |
| 1.P | Fut.1.S | - ò | + | emmo |
| 2.P | Fut.1.S | - ò | + | este |
| 3.P | Fut.1.S | - ò | + | ebbero |

# 5.I.2 Homogeneous clusters 1

| Infl. type | Lex. base | Examples |
|---|---|---|
| gare | ~gare | pagare |
| giare | ~giare | mangiare |
| **acére** | **~acére** | **giacére, piacére, tacére** |
| **dere_s_s** | **~adere** | **radere, evadere** |
| **ndere_s_s** | **~cendere** | **accendere, scendere** |
| **ndere_s_s** | **~rendere** | **rendere, prendere** |
| **dere_s_s** | **~odere** | **esplodere, rodere** |
| **dere_s_s** | **~rdere** | **ardere, perdere, mordere** |
| **dere_s_s** | **~udere** | **chiudere, accludere, alludere, prudere, intrudere** |
| **ggere_ss_tt** | **~eggere** | **leggere, reggere, proteggere** |
| **gere_s_s** | **~ulgere** | **fulgere, indulgere** |
| **gere_s_t** | **~angere** | **frangere, piangere, tangere** |
| **gere_s_t** | **~pingere** | **pingere, spingere** |
| **gere_s_t** | **~ungere** | **ungere, fungere, giungere, mungere, pungere** |
| **gere_s_t** | **~orgere** | **scorgere, porgere, sorgere** |
| **gere_s_t** | **~urgere** | **assurgere, urgere** |
| **gliere_s_t** | **~gliere** | **scegliere, cogliere, scogliere, togliere** |
| ere | ~emere | gemere, fremere, premere |
| ere | ~etere | metere, ripetere |
| **ire** | **~ucire** | **cucire, sdrucire** |

# 5.I.2 Homogeneous clusters 2

## Homogeneous clusters of verbs with infinitive ending in *-ére*

***~acére**: giacere (giaccio giaci giacciamo giacqui giaciuto), piacere, tacere*

## Homogeneous clusters of verbs with infinitive ending in *-ere*

***~adere**: radere (rado radi radiamo rasi raso), evadere*

***~cendere**: accendere (accendo accendi accendiamo accesi acceso), scendere*

***~rendere**: rendere (rendo rendi rendiamo resi reso), prendere*

***~odere**: esplodere (esplodo esplodi esplodiamo esplosi esploso), rodere*

***~rdere**: ardere (ardo ardi ardiamo arsi arso), perdere* (also reg.)*, mordere*

***~udere**: chiudere (chiudo chiudi chiudiamo chiusi chiuso), accludere, alludere, prudere* (def.)*, intrudere*

***~eggere**: leggere (leggo leggi leggiamo lessi letto), reggere, proteggere*

***~ulgere:** rifulgere (rifulgo rifulgi rifulgiamo rifulsi rifulso), indulgere* (def.)

# 5.I.2 Homogeneous clusters 3

***~angere****: frangere (frango frangi frangiamo fransi franto), piangere, tangere* (def.)

***~pingere****: pingere (pingo pingi pingiamo pinsi pinto), spingere*

***~ingere*** (nearly homogeneous)*:* to *-pingere* add *cingere, fingere, mingere, tingere, attingere;* exception *stringere (stretto)*

***~ungere****: ungere (ungo ungi ungiamo unsi unto), fungere, giungere, mungere, pungere*

***~orgere****: scorgere (scorgo scorgi scorgiamo scorsi scorto), porgere, sorgere*

***~urgere****: assurgere (assurgo assurgi assurgiamo assursi assurto), urgere* (def.)

***~gliere****: scegliere (scelgo scegli scegliamo scelsi scelto), cogliere, scogliere, togliere*

***~emere*** (reg)*: gemere (gemo gemi gemiamo gemei/etti, gemuto), fremere, premere*

***~etere*** (reg)*: metere (meto meti metiamo metei/etti, metuto), ripetere*

## Homogeneous clusters of verbs with infinitive ending in *-ire*

***~ucire****: cucire (cucio cuci cuciamo cucii cucito), sdrucire*

***~uggire*** (to be discussed)*: fuggire (fuggo fuggi fuggiamo fuggii fuggito), muggire* (also reg.)*, ruggire* (also reg.)

***~offrire*** (only compound verbs of the same basic verb)*: offrire (offro offri offriamo offrii/offersi offerto), soffrire*

***~prire*** (only compound verbs of the same basic verb)*: aprire (apro apri apriamo aprii/apersi aperto), coprire*

# 5.I.2 Homogeneous clusters 4

## Phonetic-orthographic specialties of verbs with infinitive ending in *-are*

### Verbs in *-care*:

***c/ch* alternation which conserves pronunciation.**
**This cluster is nearly homogeneous with the exception of the *o/uo* vowel alternation regarding *giocare*, but contains many trivial homogeneous sub-clusters.**

### Verbs in *-gare*:

***g/gh* alternation which conserves pronunciation.**
**This cluster is homogeneous.**

### Verbs in *-iare*:

**Using reverse sorting, it is nearly impossible to distinguish the verbs with syllabic *i* in present tense (type *desìo*) from those with non-syllabic *i* (type *copio*).**
**Only the cluster *-giare* is homogeneous (non-syllabic *i*).**
**Only the cluster *-ciare* is nearly homogeneous (non-syllabic *i*), with the exception *sciare* (syllabic *i*).**
**[For numeric evaluations, the verbs with syllabic *i* are counted as irreg.]**

# 5.I.3 Inhomogeneous clusters

## Conjugation with infinitive in *-ere*

***~idere*** (nearly homogeneous)*: decidere (decisi, deciso), elidere, ridere, intridere, assidersi, conquidere, dividere*
**exception:** *stridere* (reg)

***~endere*** (not homogeneous)*: prendere (presi, preso)* etc.
**exceptions:** *pendere* (reg), *vendere* (reg) etc.

***~ondere*** (not homogeneous)*: rispondere (risposi, risposto)* etc.
**exceptions:** *fondere (fusi, fuso), contundere*

***~ngere*** (nearly homogeneous)*: spengere (spensi, spento), ~angere, ~ingere ~ungere*
**exception:** *stringere (strinsi, stretto)*

***~ingere*** (nearly homogeneous)*: cingere (cinsi, cinto), fingere, mingere, pingere, spingere, tingere, attingere*
**exception:** *stringere (strinsi, stretto)*

***~ergere*** (nearly homogeneous)*: emergere (emersi, emerso), aspergere, tergere, convergere*
**exception:** *ergere (ersi, erto)*

**etc.**

## Conjugation with infinitive in *-ire*

***~uggire*** (nearly homogeneous)*: fuggire (fuggo)*
**exceptions:** *muggire* (also reg), *ruggire* (also reg)

**etc.**

# 5.I.4 Basic clusters

| Cluster information | Basic cluster: morphological property (infl. type) | Basic cluster: alphabetic property (lexical base) | Examples | Exceptions |
|---|---|---|---|---|
| basC | are | ~are | amare | fare |
| basC2 | care | ~care | marcare | giocare |
| basC2 | iare | ~iare | studiare | espiare |
| basC3 | ciare | ~ciare | baciare | sciare |
| **basC** | **dere_s_s** | **~dere** | **radere** | **cedere** |
| **basC2** | **dere_s_s** | **~idere** | **decidere** | **stridere** |
| **basC2** | **ndere_s_s** | **~ndere** | **accendere** | **pendere** |
| **basC3** | **ndere_s_s** | **~endere** | **prendere** | **pendere** |
| **basC3** | **ndere_s_st** | **~ondere** | **rispondere** | **fondere** |
| **basC** | **gere_s_t** | **~gere** | **volgere** | **stringere** |
| **basC2** | **ggere_ss_tt** | **~ggere** | **leggere** | **affiggere** |
| **basC2** | **gere_ss_tt** | **~igere** | **diligere** | **esigere** |
| **basC2** | **gere_s_s** | **~lgere** | **fulgere** | **volgere** |
| **basC2** | **gere_s_t** | **~ngere** | **spengere** | **stringere** |
| **basC3** | **gere_s_t** | **~ingere** | **cingere** | **stringere** |
| **basC2** | **gere_s_t** | **~rgere** | **scorgere** | **emergere** |
| **basC3** | **gere_s_s** | **~ergere** | **emergere** | **ergere** |
| basC | ire(sc) | ~ire | finire | uscire |
| basC2 | ire | ~uggire | fuggire | muggire |

**One can dispute about the cluster written in orange.**

# 5.I.5 Cluster tree

| Level 1 | Level 2 | Level 3 | Level 4 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|---|
| L1 | | | | basC | are | ~are | amare | fare |
| | L2 | | | basC2 | care | ~care | marcare | giocare |
| | L2 | | | homC | gare | ~gare | pagare | -- |
| | L2 | | | basC2 | iare | ~iare | studiare | espiare |
| | | L3 | | basC3 | ciare | ~ciare | baciare | sciare |
| | | L3 | | homC | giare | ~giare | mangiare | -- |
| **L1** | | | | **homC** | **acére** | **~acére** | **piacére** | **--** |
| **L1** | | | | **basC** | **dere_s_s** | **~dere** | **radere** | **cedere** |
| | **L2** | | | **homC** | **dere_s_s** | **~adere** | **radere** | **--** |
| | **L2** | | | **basC2** | **dere_s_s** | **~idere** | **decidere** | **stridere** |
| | **L2** | | | **basC2** | **ndere_s_s** | **~ndere** | **accendere** | **pendere** |
| | | **L3** | | **basC3** | **ndere_s_s** | **~endere** | **prendere** | **pendere** |
| | | | **L4** | **homC** | **ndere_s_s** | **~cendere** | **accendere** | **--** |
| | | | **L4** | **homC** | **ndere_s_s** | **~rendere** | **rendere** | **--** |
| | | **L3** | | **basC3** | **ndere_s_st** | **~ondere** | **rispondere** | **fondere** |
| | **L2** | | | **homC** | **dere_s_s** | **~odere** | **esplodere** | **--** |
| | **L2** | | | **homC** | **dere_s_s** | **~rdere** | **ardere** | **--** |
| | **L2** | | | **homC** | **dere_s_s** | **~udere** | **chiudere** | **--** |
| **L1** | | | | **basC** | **gere_s_t** | **~gere** | **volgere** | **stringere** |
| | **L2** | | | **basC2** | **ggere_ss_tt** | **~ggere** | **leggere** | **affiggere** |
| | | **L3** | | **homC** | **ggere_ss_tt** | **~eggere** | **leggere** | **--** |
| | **L2** | | | **basC2** | **gere_ss_tt** | **~igere** | **diligere** | **esigere** |
| | **L2** | | | **basC2** | **gere_s_s** | **~lgere** | **fulgere** | **volgere** |
| | | **L3** | | **homC** | **gere_s_s** | **~ulgere** | **fulgere** | **--** |
| | **L2** | | | **basC2** | **gere_s_t** | **~ngere** | **spengere** | **stringere** |
| | | **L3** | | **homC** | **gere_s_t** | **~angere** | **frangere** | **--** |
| | | **L3** | | **basC3** | **gere_s_t** | **~ingere** | **cingere** | **stringere** |
| | | | **L4** | **homC** | **gere_s_t** | **~pingere** | **pingere** | **--** |
| | | **L3** | | **homC** | **gere_s_t** | **~ungere** | **giungere** | **--** |
| | **L2** | | | **basC2** | **gere_s_t** | **~rgere** | **scorgere** | **emergere** |
| | | **L3** | | **basC3** | **gere_s_s** | **~ergere** | **emergere** | **ergere** |
| | | **L3** | | **homC** | **gere_s_t** | **~orgere** | **scorgere** | **--** |
| | | **L3** | | **homC** | **gere_s_t** | **~urgere** | **assurgere** | **--** |
| **L1** | | | | **homC** | **gliere_s_t** | **~gliere** | **scegliere** | **--** |
| L1 | | | | homC | ere | ~emere | premere | -- |
| L1 | | | | homC | ere | ~etere | ripetere | -- |

| Level 1 | Level 2 | Level 3 | Level 4 | Cluster type | Inflection type | Cluster (an. / synth. lex. base) | Examples | Exceptions |
|---|---|---|---|---|---|---|---|---|
| L1 | | | | basC | ire(sc) | ~ire | finire | uscire |
| | **L2** | | | **homC** | **ire** | **~ucire** | **cucire** | **--** |
| | L2 | | | basC2 | ire | ~uggire | fuggire | muggire |

# 5.I.6 Part of the lexeme register

## see Excel table

# 5.I.7 Result

**There are a regular, two mixed and an irregular conjugation class.**
**Using reverse sorting, they can be distinguished exactly**
**if one marks the accent of the verbs in *~ére* und *~ere*:**

**Verbs in *~are*: mostly regular verbs (productive).**

**Verbs in *~ire*: regular (stem-extending, productive) and irregular (not stem-extending, not productive);**
**cannot be completely distinguished using the reverse view.**

**Verbs in *~ére*: irregular verbs (not productive).**

**Verbs in *~ere* (not productive): few regular verbs (type *vendere*) and many irregular verbs;**
**cannot be completely distinguished using the reverse view.**
**There are only the clusters *~emere*, *~etere* which are homogeneous and contain regular verbs.**

**There are numerous starting points for assumptions of analogy on the basis of reverse similarity.**
**About one fourth (50/200) of the irregular verbs**
**are elements of homogeneous clusters of reversely similar verbs with the average size 2 ½ (50/18).**

**The investigation is based on Holl 1988: 231-246.**

# 6. References

## General references


Alpar, Paul; Niedereichholz, Joachim: *Data Mining im praktischen Einsatz*. Braunschweig / Wiesbaden 2000.

Ester, Martin; Sander, Jörg: *Knowledge discovery in databases. Techniken und Anwendungen*. Berlin 2000.

Friederici, Angela: "Wie das Gehirn zur Sprache kommt". *Spektrum der Wissenschaft* 2010: 1, 66-71.

Holl, Alfred; Suljić, Ivan: *Rückläufiges Wörterbuch zur kroatischen Verbalmorphologie. Aufbereitung mit Datenanalyseverfahren der Informatik (Data Mining)*. Regensburg: Roderer 2010 [= Studia et exempla linguistica et philologica, Series V: Lexica, Tom. 6].

Holl, Alfred; Zimnik, Gordon: "Data Mining und natürlichsprachliche Verbalmorphologien". [= Schriftenreihe der Georg-Simon-Ohm-Hochschule Nürnberg 43(2008) 1-54]. Nürnberg: Ohm-Hochschule 2009.

Holl, Alfred; Maroldo, Sara; Urban Reinhard: *The inflectional morphologies of the Swedish noun, the Swedish verb and the English verb. Reverse dictionaries based upon data mining methods*. Växjö 2007 [= Mathematical modelling in physics, engineering and cognitive sciences, vol. 12].

Holl, Alfred; Pavlidis, Stilianos; Urban, Reinhard: *Rückläufiges Wörterbuch zur alt- und neugriechischen Verbalmorphologie. Aufbereitung mit Datenanalyse-Verfahren der Informatik (Data Mining)*. Regensburg 2006 [= Studia et exempla linguistica et philologica, Series V: Lexica, Tom. 5].

Holl, Alfred; Behrschmidt, André; Kühn, Alexander: *Rückläufige Register zur russischen und deutschen Verbalmorphologie. Aufbereitung mit Datenanalyse-Verfahren der Informatik (Data Mining)*. Regensburg 2004 [= Studia et exempla linguistica et philologica, Series V: Lexica, Tom. 4].

Holl, Alfred: "Datenanalyseverfahren der Informatik (Data Mining) als Grundlage einer didaktischen Darstellung der französischen Verbalmorphologie". In: Bernhard, Gerald; Kattenbusch, Dieter; Stein, Peter (ed.): *Namen und Wörter. Festschrift Josef Felixberger zum 65. Geburtstag*. Regensburg 2003, 107-119.

Holl, Alfred: "Licht und Schatten von Analogieschlüssen auf der Basis rückläufiger Ähnlichkeit in der Verbalmorphologie romanischer und germanischer Sprachen". In: Heinemann, Sabine; Bernhard, Gerald; Kattenbusch, Dieter (ed.): *Roma et Romania. Festschrift Gerhard Ernst zum 65. Geburtstag*. Tübingen 2002, 152-167.

Holl, Alfred: "The inflectional morphology of the Swedish verb with respect to reverse order: analogy, pattern verbs and their key forms". *Arkiv för nordisk filologi* 116 (2001), 193-220.

Holl, Alfred: *Romanische Verbalmorphologie und relationentheoretische mathematische Linguistik. Axiomatisierung und algorithmische Anwendung des klassischen Wort und Paradigma-Modells*. Tübingen 1988 [= Linguistische Arbeiten 216].

Kruse, Rudolf: Borgelt, Christian: "Suche im Datendschungel". *Spektrum der Wissenschaft* 2002: 11, 80-81.

**Selected examples of verb registers and reverse dictionaries**

Alinei, Mario L.: *Dizionario inverso italiano*. Den Haag 1962.

Edmonds, David: *The Oxford reverse dictionary*. Oxford 2002.

Einberger, Angela: *Langenscheidt Verb-Tabellen Englisch.* Berlin 2005, 2000.

Elia, Pietro: *I verbi italiani per gli stranieri*. Mailand [1955] 1983.

Goulding, Sylvia: *Englisch. Verben. Basiswissen. Eine leicht verständliche Beschreibung des englischen Verbsystems.* Princeton NJ $^{5}$1998.

Juilland, Alphonse: *Dictionnaire inverse de la lange française*. Den Haag 1965.

Kempgen, Sebastian: *Grammatik der russischen Verben.* Wiesbaden 1989.

Langendorf, Dieter: *L'art de conjuguer. Le Nouveau Bescherelle: Dictionnaire de douze mille verbes*. Frankfurt 1986.

Lehnert, Martin: *Rückläufiges Wörterbuch der englischen Gegenwartssprache.* Leipzig $^{2}$1973.

Mateo, Francis: *El arte de conjugar en espanol. Diccionario de 12 000 verbos*. *Collection Bescherelle.* Paris 1995.

Mater, Erich: *Rückläufiges Wörterbuch der deutschen Gegenwartssprache.* Leipzig 1965.

Muthmann, Gustav: *Reverse English dictionary based on phonological and morphological principles*. Berlin 2002, 1999.

Muthmann, Gustav: *Rückläufiges deutsches Wörterbuch*. Tübingen 2001.

Scott, Samantha: *Pons Verbtabellen Englisch*. Stuttgart 2001.

Sleumer, Albert: *Die unregelmäßigen Zeitwörter der goldenen und silbernen Latinität in ihren einfachen und zusammengesetzten Formen*. Bonn $^{2}$1962.

Stahl, Fred A.; Scavnicky, Gary E. A.: *A reverse dictionary of the Spanish language*. Urbana 1973.

Weermann, Eva Maria: *Pons Verbtabellen Deutsch.* Stuttgart 2001.

Zaliznjak, Andrej A.: *Grammatičeskij slovar' russkogo jazyka: slovoizmenenie*. Moskau $^{4}$2003 = $^{3}$1987 = $^{1}$1977 [= Зализняк, Андрей А.: *Грамматический словарь русского языка: словоизменение*].