

The database of Estonian Word Families

Ülle Viks, Silvi Vare, Heete Sahkai
Institute of the Estonian Language

Outline

1. Background

- What is a word family
- The word families method

2. Data

3. Design: editing, query, Web interface

4. Applications

Word family

A word family (WF) is the set of all the words in the vocabulary of a language that contain a common stem morpheme:

- **aed** 'garden n.'
- **aednik** 'gardener'
- **aedmaasikas** 'garden strawberry'
- **aeda** pidama 'garden v.'
- **aiapidaja** 'gardener'



Word family

The WF is introduced by the simplex word that represents the common stem – the head of the family:

- AED
- aednik
- aedmaasikas
- aeda pidama
- aiapidaja

Word family

The words in the WF – the family members – are analyzed into immediate constituents and are assigned a word formation type:

- **aed=nik** ‘garden=noun suffix’
- **aed+maasikas** ‘garden+strawberry’
- **aeda pidama** ‘garden.partitive keep’
- **aia+pida=ja** ‘garden.genitive+keep+noun suffix’

Word family

The words in the WF – the family members – are organized hierarchically according to mutual word formational relations: each word is preceded by its base word and followed in turn by the derivations and compounds that are based on it:

AED ‘garden’

lasteaed ‘child.gen.pl+garden’ “kindergarten”

lasteaednik ‘kindergarten=noun suffix’

“kindergarten teacher”

Word family

- ELA#MA

- ela=mu

- kahe+pere+ela|mu

- ela=nik

- ela|nik=kond

- el=u

- el|u=s

- el|u=tu

- abi+el|u

- abi α ellu astu#ma

- abi α ell|u ξ #ma

- abi α ell|u=mine

- abi α ell|u|mis+ette α pane|k



The word families method

- Consists in organizing the entire vocabulary of a language into word families
- A method for structuring the vocabulary of a language
- A way of representing the word formation of a language
- The method used in the compilation of word formation dictionaries
- Consists in the word formation analysis of all the words of a language
- Presupposes a detailed description of word formation in the language

Principal word formation dictionaries

- Augst, Gerhard 1998.
Wortfamilienwörterbuch der deutschen Gegenwartssprache. Tübingen: Max Niemeyer Verlag.
- Splett, Jochen 2009. **Deutsches Wortfamilienwörterbuch.** Analyse der Wortfamilienstrukturen der deutschen Gegenwartssprache, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes. Berlin/New York: de Gruyter.
- Tikhonov, A., N. 1985. Slovoobrazovatel'nyj slovar' russkogo jazyka I–II. Moskva: Russkii jazyk.
- ...



The WF method as the design principle of an electronic database

- A new type of linguistic resource
- Greatly improved access to word formation data and description
- A wide range of potential applications

Data

- The inventory of words is based on the latest large general dictionaries of Estonian
- The word formation analysis is based on the descriptive grammar of Estonian and subsequent research into Estonian word formation
- 8880 word families
- 192 000 items in total

Units of the macrostructure of the database: the word family

aed subst.

- [P_TUL]
 - aed=ik subst. *väike aed*
 - aed=nik subst.
 - maa|stiku+aed|nik subst. (*tegeleb maastiku kujundamisega*)
- [P_LS1]
 - botaanika+aed subst.
 - ema+aed subst. *aiand.* (*kust võetakse seemneid, pook- ja pistoksi*)
 - las#te+aed subst.
 - las#te α aed=nik subst. *lasteaiakasvataja*
 - las#te α aia+kasva|taja subst.
 - las#te α aia+laps subst.
- [P_LS2]
 - aed+maasikas subst.
 - aed α maasika+kee|d|is subst.
 - aia+maja subst.
- [P_YH2]
 - aeda pida#ma
 - aia+pida=ja subst.
 - aia+pida=mine subst.

Units of the macrostructure of the database: the word family

- The word family is introduced by the head of the word family (a simplex word) and constituted of the family members.
- The family members are organized hierarchically by step of formation.
- The maximal number of steps found in the database is seven.

Units of the macrostructure of the database: the word family

- On the first level of the hierarchy, the head is followed by all the words based on it – the first-step formations. For clarity of presentation, the first-step formations are divided into separate blocks according to their word formation kind: derivatives (P_TUL), compounds by the second constituent (P_LS1), compounds by the first constituent (P_LS2), verbal expressions by the second constituent, and verbal expressions by the first constituent (P_YH2).
- Each first step formation is again followed by the eventual second-step formations, that is the words that are in turn based on it, and so forth.

Units of the macrostructure: family members

- AED ‘garden’
- aed=nik ‘garden=noun suffix’
 - maastiku+aed|nik
‘landscape.gen+garden|noun suffix’
- aed+maasikas ‘garden+strawberry’
- aeda pidama ‘garden.partitive keep’
 - aia+pida=ja
‘garden.genitive+keep+noun suffix’

The units of the microstructure

Each head of family and each family member has its own microstructure, separate fields for representing grammatical and lexical information about them:

- Homonym number
- Part-of-speech
- Definition
- Subject label
- Usage label
- Context
- ...

Design

- Embedded in the dictionary management system EELex
- Based on a specially designed XML schema that follows the hierarchical structure of word families
- Provided with a Web interface

The dictionary management system EELEX

- A web-based toolset for dictionary writing and management
- Stores universal reusable databases encoded in a standard XML format
- Provides tools for editing, query, layout design

EELex editing window

- The editing window is divided into the editing pane and the layout pane, which are mutually connected by click.
- In the editing pane, data can be edited both in table form and in the XML code.

EELex editing window: table view

The screenshot shows the EELex editing window in table view. The browser address bar displays `http://eelex.dyn.eki.ee/_shs/art_sp_cg`. The search bar contains the word `jala`. The main editing area shows a table with the following structure:

Toimetamisala	XML	Tabelina	Vaadena
↓	ÖS-grupp		
PLOKKIDE TSOON			
	Tuletusplokk		↓
	Taane		↓
	Liikme var-grupp		↓
	pereliige	[-]	↓
	sõnaliik	[+]	↓
	Tähendusgrupp		↓
	tähendus		↓
	ÖS-grupp		
	Taane		↓
	Liikme var-grupp		↓
	pereliige	[+]	↓
	Tähendusgrupp		↓
	valdkond		↓
	ÖS-grupp		

The right pane shows the morphological analysis for `jala=m` (SUBST. *olenditel, asjadel*; *endisaegne pikkusmõõt, u. 30 cm*):

- [P_TUL]
 - jala=m SUBST. *mäe v. künkajalg; ehitise alumine osa*
 - mandri+jala|m GEOGR.
 - mäe+jala|m
 - trep+jala|m
 - jala=nd SUBST. *aparaadi vm. jalgosa, statiiv*
 - jala=s SUBST. (*sõidukil; hällil vm.*)
 - hälli+jala|s
 - kelgu+jala|s
 - ree+jala|s
 - saani+jala|s
 - jala|s+häll
 - jala|s+kulgmik
 - jala|s+sõiduk
 - jala|se+painjik
 - jala|se+pakk
 - jala=ts SUBST. *jalanõu*

EELex editing window: XML view

The screenshot shows the EELex XML editing window for the word "jalg". The interface includes a search bar with "jalg" entered, a toolbar, and a main editing area. The XML structure is as follows:

```
<p:A>
  <p:P>
    <p:mg>
      <p:mag>
        <p:m p:O="jalg" p:mm="">jalg
      <p:sl>subst.
      <p:dg>
        <p:d>(olenditel, asjadel)
      <p:dg>
        <p:d>endisaegne pikkusmõõt, u. 30 cm
    <p:plokid>
      <p:p_tul>
        <p:t>
          <p:lag>
            <p:ml p:mm="">jala=m
          <p:sl>subst.
          <p:dg>
            <p:d>mäe v. künkajalg; ehitise alumine osa
          <p:t>
            <p:lag>
```

The right-hand pane displays the dictionary entry for "jalg":

jalg SUBST. (olenditel, asjadel); endisaegne pikkusmõõt, u. 30 cm

- [P_TUL]
 - jala=m SUBST. mäe v. künkajalg; ehitise alumine osa
 - mandri+jala|*m* GEOGR.
 - mäe+jala|*m*
 - trep+jala|*m*
 - jala=nd SUBST. aparaadi *vm.* jalgosa, statiiv
 - jala=s SUBST. (sõidukil; hällil *vm.*)
 - hälli+jala|*s*
 - kelgu+jala|*s*
 - ree+jala|*s*
 - saani+jala|*s*
 - jala|*s*+häll
 - jala|*s*+kulgmik
 - jala|*s*+sõiduk
 - jala|*s*e+painjik
 - jala|*s*e+pakk
 - jala=ts SUBST. jalanõu

Editing

- For the hierarchical DEWF, important editing functions are the adding, deleting and moving of whole structural groups (blocks and family members).
- Another important editing function is bulk corrections because a large number of words occur in two or more word families and this function permits to modify all these occurrences at once.

Editing: block moving

The screenshot shows the 'SõnaPered: 'YViks'' web application in Internet Explorer. The search bar contains 'märksõna' and the search term 'jalg'. The search results are displayed in a tree view on the left and a list on the right. The tree view shows a hierarchy of words, with 'jalg' selected. The list on the right shows the following words and their meanings:

- jalutuz 2. jalutu inimene
- jalutu=lt \maas lamama\
- jalutu=us 1. SUBST. *jalutu olek*
- o jalu=ta#ma
 - jalutu=ele#ma *jalutlema*
 - jalutu|tel=u
 - jalutu=le#ma *edasi-tagasi v. ringi jalutama*
 - jalutu=ja
 - jalutu=mine
 - jalutu=us 2. SUBST. *jalutamine*
 - jalutu|tus+kepp
 - jalutu|tus+käik
 - jalutu|tus+rihm
 - jalutu|tus+ruum
 - o jala# ADV. *jalgsi*
 - jala=sta#ma *SÕJ. jalameheks tegema*
 - jala=stu#ma *hobuse seljast v. lahingumasinalt maha tulema*
 - jala|stu=mine
 - jala+matk
 - jala+mees *jalakäija; AJ. teoline, kes töötas*

The status bar at the bottom indicates: 'Uisaldusväärased kohad' and '100%'.

Editing: bulk corrections

http://eelex.eki.ee/ - SP_ - <> - Windows Internet Explorer

Otsitav piirkond: [1,∞] - <p.A> - artikkel

Otsitav element: [1,1] - <p.ml> - pereliige

lisa element (p.t) grupis

A - K tõstatundetu
 L - Q märkidega
 R - Y globaalselt

Üksikud
 +
 -

Otsi: [0,∞] <p.v> valdkond

Rakenda

Rak. puud.

Tühista

Võrdle taandeid

sport

<input type="checkbox"/>	kerge	<p:t>	• kerge=jõu stiku+võist use#d SUBST. <i>SPORT</i> kerge □ jõustik		kerge□jõu stiku+võist use#d
<input checked="" type="checkbox"/>	maa	<p:t>	• maa=ilma=meistri+võist use#d SUBST. maa□ilma□meister	+ <p.dg/p.v> = "sport"	maa□ilma□meister+võist use#d
<input checked="" type="checkbox"/>	male	<p:t>	• male+meistri=võist use#d SUBST. male+meister	+ <p.dg/p.v> = "sport"	male+meistri□võist use#d
<input checked="" type="checkbox"/>	meister	<p:t>	• meistri+võist use#d SUBST. .meister+ võistused	+ <p.dg/p.v> = "sport"	meister+võist use#d
<input checked="" type="checkbox"/>	meister	<p:t>	• jalg=palli-meistri=võist use#d SUBST. jalg□pall	+ <p.dg/p.v> = "sport"	jalg□palli-meistri□võist use#d
<input checked="" type="checkbox"/>	meister	<p:t>	• kerge=jõu stiku-meistri=võist use#d SUBST. .meister□ võistused	+ <p.dg/p.v> = "sport"	kerge□jõu stiku-meistri□võist use#d
<input checked="" type="checkbox"/>	meister	<p:t>	• male+meistri=võist use#d SUBST. male+meister	+ <p.dg/p.v> = "sport"	male+meistri□võist use#d
<input checked="" type="checkbox"/>	meister	<p:t>	• võrk=palli-meistri=võist use#d SUBST. võrk□pall	+ <p.dg/p.v> = "sport"	võrk□palli-meistri□võist use#d
<input type="checkbox"/>	miss	<p:t>	• missi+võist use#d SUBST.		missi+võist use#d
<input checked="" type="checkbox"/>	pall 1	<p:t>	• jalg=palli+meistri=võist use#d SUBST. jalg□pall	+ <p.dg/p.v> = "sport"	jalg□palli+meistri□võist use#d
<input type="checkbox"/>	" 1	<p:t>	• võrk=palli+meistri=võist use#d SUBST. võrk□pall	+ <p.dg/p.v> = "sport"	võrk□palli+meistri□võist use#d

Käivita op.

Otsing: 0m, 16s; 34 objekti, 19 art.

Query

- The EELEX software permits to conduct structure based queries by every labelled group, element and attribute.
- The search results can be sorted in different ways: each column can be sorted in increasing, decreasing and reverse order (i.e. by the final letters of words).

Query: words with a usage label

SõnaPered: 'Yviks' - Windows Internet Explorer

http://eelex.dyn.eki.ee/_shs/art_sp_.cgi

Fail Redigeeri Vaade Lemmikud Iõõriistad Spikker

Google Search Bookmarks Check AutoFill Sign In

Lemmikud Pakutud saidid

SõnaPered: 'Yviks'

Kõide: Kõik kõited

stiil

Otsi

Päring: [stiil (./p:s) ['*' (↔: 1)]], tt-u, m-ta, glob.; 4218 leidu, 2023 artiklit.

Art-jnr	Kd.	Märksõna(d)	Leid	K.	T.	Pt.
1	I	<m> aabits	<p:s> •NALJ.	EKI	SVare	
2	I	<m> Aadam	<p:s> •NALJ.	EKI	SVare	
2	I	<m> Aadam	<p:s> •PILTL.	EKI	SVare	
2	I	<m> Aadam	<p:s> •PILTL.	EKI	SVare	
2	I	<m> Aadam	<p:s> •PILTL.	EKI	SVare	
2	I	<m> Aadam	<p:s> •PILTL.	EKI	SVare	
3	I	<m> aader	<p:s> •VNM.	EKI	SVare	
3	I	<m> aader	<p:s> •PILTL.	EKI	SVare	
4	I	<m> aadress	<p:s> •VNM.	EKI	SVare	
5	I	<m> aare	<p:s> •PILTL.	EKI	SVare	
6	I	<m> aaria , aarja	<p:s> •VNM.	EKI	SVare	
7	I	<m> aasta	<p:s> •KÕNEK.	EKI	SVare	
7	I	<m> aasta	<p:s> •KÕNEK.	EKI	SVare	
7	I	<m> aasta	<p:s> •KÕNEK.	EKI	SVare	

[stiil (./p:s) ['*' (↔: 1)]], tt-u, m-ta, glob.: leiti 2023 artiklit; (0m, 5s)

Usaldusväärsed kohad 100%

Query: part-of-speech = adv. (reverse order by final letters)

SõnaPered: 'YViks' - Windows Internet Explorer

http://eelex.dyn.eki.ee/_shs/art_sp_...cgi

Fail Redigeeri Vaade Lemmikud Tööriistad Spikker

Google Search Bookmarks Check AutoFill Sign In

SõnaPered: 'YViks'

Kõide: 1. köide A - K

sõnaliik adv.

Otsi

Päring: [sõnaliik (p:P/p:mg/p:sl) ['adv\.' (↔: 5)], tt-u, m-ga, lok.; 122 leidu, 103 artiklit.

Art-jnr	Kd.	Märksõna(d)	Leid	K.	T.	Pt.
27	I	<m> jaa_jah_jaa_jah_jaa	<p:sl> •ADV.	SVare	SVare	SVare
31	I	<m> juba	<p:sl> •ADV.	EKI	SVare	SVare
12	I	<m> ega	<p:sl> •ADV.	EKI	SVare	SVare
18	I	<m> hilja	<p:sl> •ADV.	EKI	SVare	SVare
25	I	<m> ikka	<p:sl> •ADV.	EKI	SVare	SVare
22	I	<m> huka#s_hukka	<p:sl> •ADV.	EKI	SVare	SVare
22	I	<m> huka#s_hukka	<p:sl> •ADV.	EKI	SVare	SVare
38	I	<m> ka	<p:sl> •ADV.	EKI	SVare	
3	I	<m> ala	<p:sl> •ADV.	EKI	SVare	
1	I	<m> aina	<p:sl> •ADV.	EKI	SVare	
48	I	<m> kaua	<p:sl> •ADV.	EKI	SVare	SVare
2	I	<m> aiva	<p:sl> •ADV.	EKI	SVare	
64	I	<m> kohe	<p:sl> •ADV.	EKI	SVare	SVare
26	I	<m> irevil_irevi#le	<p:sl> •ADV.	EKI	SVare	SVare
41	I	<m> kalkvel_kalkvele	<p:sl> •ADV.	EKI	SVare	SVare
41	I	<m> kalkvel_kalkvele	<p:sl> •ADV.	EKI	SVare	SVare

[sõnaliik (p:P/p:mg/p:sl) ['adv\.' (↔: 5)], tt-u, m-ga, lok.: leiti 103 artiklit; (0m, 2s)

Usaldusväärased kohad 100%

Query: the block zone is empty

The screenshot shows a search result page for the query 'plokkide tsoon'. The search criteria are: 'plokkide tsoon', '=NULL', and './text()'. The results show 332 articles found. The table below lists the first 10 results.

Art-jnr	Kd.	Märksõna(d)	Leid	K.	T.	Pt.
1	I	<m> aamen	<p:A> •aamen =st="m0" =mm="" •SUBST. •RELIG. • <i>kinnitussõna palve, jutluse vms. lõpul</i>	EKI	SVare	
2	I	<m> aanispits	<p:A> •aanispits =mm="" •pits =i="2"	SVare	SVare	
3	I	<m> aara	<p:A> •aara =st="m0" =mm="" •SUBST. •ZOOLOG. •(<i>papagoi</i>)	EKI	SVare	
4	I	<m> abee	<p:A> •abee =st="m0" =mm="" •SUBST. •RELIG. • <i>katoliku alamvaimulik Prantsusmaal</i>	EKI	SVare	
5	I	<m> adjöö	<p:A> •adjöö =st="m0" =mm="" •INTERJ. •VNM. • <i>hüvasti</i>	EKI	SVare	
6	I	<m> ahm	<p:A> •ahm =st="m0" =mm="" •SUBST. •ZOOLOG. • <i>kaljukass</i>	EKI	SVare	
7	I	<m> ai	<p:A> •ai =i="2" =st="m0" =mm="" •SUBST. •ETN. • <i>seeliku äärispael</i>	EKI	SVare	
8	I	<m> ain	<p:A> •ain =st="m0" =mm="" •SUBST. •KAL. • <i>võrgu kinnitusaas</i>	EKI	SVare	
9	I	<m> aituma, aitüma	<p:A> •aituma =st="m0" =mm="" •aitüma =st="m0" =mm="" •aitäh	EKI	SVare	
10	I	<m> aiva	<p:A> •aiva =i="1" =st="m0" =mm="" •ADV. • <i>aina, ikka</i>	EKI	SVare	

[plokkide tsoon (./p:plokid) ['=NULL' (↔: 5)], tt-u, m-ta, glob.: leiti 332 artiklit; (0m, 2s)

Usaldusväärsed kohad 100%

Web interface

- The resources completed in EELex are made available through the Web as free public resources.
- The Web interface supports structure-based querying.

Web interface: structure-based query

[SP] "Eesti keele sõnapered" - Windows Internet Explorer

http://www.eki.ee/dict/sp/index.cgi?Q=sti&F=5X5

Fail Redigeeri Vaade Lemmikud Tööriistad Spikker

Lemmikud Tallinn Eurolines Lux Expressi uhiuu... Otsi piletit HLT 2010, October 7-8, Rig... Keelevaeb EK Keeleressursid Soovitatud saidid

SõnaPered: 'Viks'

[SP] "Eesti keele sõnapered"

Tööversioon!
Kasutajate tagasiside on oodatud: Silvi Vare

Päring: sti [Otsi ja näita] artikli osas ...sufiks

- Perepea
- Pereliige
- Mall
- ...liitsõnaosa
- ...prefiks
- ...sufiks**
- ...morf. formaativ
- ...juur
- Kõigiit väljadelt

Leitud 173 artiklit

ablas ADJ. *ahne sööma v. jooma* [P_TUL] **apla=sti** ADV. *aplalt*

aeg SUBST. [P_TUL] **aeg=us** ADJ. *õigeaegne, piisava ajavaruga* **aeg|sa=sti** ADV.

agar 1. ADJ. *hakkaja, ettevõtlik; usin, virk* → **agar** 2. [P_TUL] **agara=sti** ADV.

ahne ADJ. [P_TUL] **ahne=sti** ADV.

arm 1. SUBST. *armastus; heasoovlikkus, soosing* → **arm** 2. **NB! ARMASTAMA teo- ja tegjanime**de vaja

Valmis Usaldusväärsed kohad 110%

start {C:\Dketas\... Inbox - Töö... NoteTab Lig... [SP] "Eesti k... Microsoft P... HLT2010 Vik... ET 17:30

Web interface: search results display

Word families are often extremely large and the queried item may thus be difficult to find in the whole entry.

Therefore we use a match-based display: in the initial search result, only family members containing the element that matches the search criteria are displayed, together with the family member[s] immediately preceding it in the hierarchy. The remaining part of the entry is hidden behind green plus-buttons.

In order to display the other family members on the same level of the hierarchy the user has to click on the green button.

Web interface: search results display

The screenshot shows a Windows Internet Explorer browser window displaying search results from the website <http://www.eki.ee/dict/sp/index.cgi?Q=sti&F=SXS>. The browser's address bar and menu bar are visible at the top. The main content area displays a list of words with their grammatical categories and related forms:

- häda** SUBST. [P_TUL] häda=sti ADV. *väga, ilmtingimata*
- hääras** ADJ. *tundlik ja hale, halemeelsusse kalduv* härda=sti
- hüva** ADJ. *hea, MURDES. parem, parempoolne* hüva=sti INTERJ. hüva|sti jä|t#ma hüva|sti+jä|tte hüva|sti+jä|tu+kiin hüva|sti+jä|tu+suu#d||us hüva|sti+jä|t=mine
- iga 1.** SUBST. *eluiga; vanus* → **iga 2.** iga,ve=ne ADJ. *alaline, kestev* iga,ve=sti
- ilu** SUBST. ilu=s ADJ. ilu|sa=sti ADV.
- ime** SUBST.

The browser's taskbar at the bottom shows the Start button, several open applications, and the system tray with the time 17:27.

Web interface: search results display

The screenshot shows a Windows Internet Explorer browser window displaying the search results for the word "häda" on the website www.eki.ee. The browser's address bar shows the URL: http://www.eki.ee/dict/sp/index.cgi?Q=sti&F=SXS. The search results are organized into three main sections: "häda", "härda", and "hüva".

häda SUBST. (+)
[P_TUL] (-)
häda=line 1. SUBST. *hädasojja*
häda=line 2. ADJ. *vasi\ kiireloomuline* (-)
häda|lise=lt
häda=ne ADJ. *hädine*
hädi=ne ADJ. *vilets, vaevaline, armetu* (-)
hädi|se=lt
hädi|s=us
häda=lda#ma *häda kurtma, häda kaebama* (-)
häda|lda=ja
häda|lda=mine SUBST.
häda|ld=us
häda=tse#ma *hädaldama* (-)
häda|tse=mine SUBST.
häda=sti ADV. *väga, ilmtingimata*

härda ADJ. *tundlik ja hale, halemeelsusse kalduv* (+)
[P_TUL] (-)
härda=lt
härda=sti
härda=us *tundlikkus, halemeelsus* (-)
härda|us+pisar SUBST.
härda|us+tunn|e
härda=u#ma (-)
härda|u=mus SUBST. *härduanud olek*

hüva ADJ. *hea, MURDES. parem, parempoolne* (+)
[P_TUL] (+)

Applications

Estonian is typologically an agglutinative-fusional language characterized by extensive stem variation and the abundance of formatives. The majority of Estonian vocabulary consists of derivations and compounds with possibly quite complex structure.

The needs generated by the Estonian word formation system:

- language description
- language education
- lexicography
- language technology

Applications: research

- Data for research into word formation and related areas
- The process of the compilation of the database has already given rise to studies into several problematic and less researched phenomena of Estonian word formation

Applications: language education

- A tool for learning Estonian word formation and the vocabulary of Estonian
- Permits to generate different types of learner's dictionaries of word formation
- A tool for teachers for compiling custom teaching materials

Applications: lexicography

- Helps to compile the lists of headwords of dictionaries
- Provides the word formation segmentation of the complex headwords of dictionaries
- Provides the lists of selected derivatives and compounds to be included in the entries of dictionaries

Applications: language technology

- Word formation module of automatic morphology
- Information retrieval
- Speech synthesis
- Integrated lexicon and grammar system

Thank you!