

Language varieties meet one-click dictionary

Andrea Abel & Egon Stemle
Institute for Applied Linguistics
Eurac Research, Bolzano/Bozen, Italy

andrea.abel@eurac.edu, egon.stemle@eurac.edu

Overview

- Aims
- Background
- Challenges
- (Semi)-automated opportunities
- Conclusions and outlook

Aims

- Aims of our research:
 - language observation and evaluation of trends of the local standard variety of the German language
 - consideration for future editions of the “Variantenwörterbuch des Deutschen” (Dictionary of variants of the German language) (Ammon et al. 2016) and other dictionariesthrough
 - the use of (semi-)automatic methods for the lexicographic processing of the German standard used in South Tyrol
 - the collaboration with international initiatives such as ELEXIS

A small excursus ...



Background

- Research on the German standard variety used in South Tyrol



- Based on the concept of pluricentricity of the German language (cf. Clyne 1992, Ammon 1995):
differing standard varieties used in the German speaking area (official status, taught in school, codices, etc.)
- South Tyrol as an interesting object of linguistic studies:
 - role as “national semi-centre” from a pluricentric perspective
 - marginal position within the German speaking area
 - the language contact situation

Challenges

Examples:

- data basis:
 - corpora: newspaper texts → suited to represent general language use?
 - example: meaning of “Bar” in the local standard German (“coffee shop” - “night bar”) (cf. Abel 2018)

<Begonnen hatte der Einsatz um 3.30 Uhr, als in der **Bar** „Ortler“ an der Staatsstraße in Schluderns die Alarmanlage losging. Einbrecher hatten versucht, die Eingangstür aufzubrechen.> (Dolomitenkorpus, 2001)

*(The operation of the police started at 3.30 am as in the **coffee shop** “Ortler” at the state road in Schluderns the alarm went on. Burglars had tried to break of the entrance door.)*

<Zu der Bluttat war es vor dem Eingang der „Bar Pleres“ in Matsch gekommen.> (Dolomitenkorpus, 2001)

(The bloody deed happened at the entrance of the “Coffee shop Pleres” in Matsch.)

Challenges

Examples:

- item classes:
 - focus on the occurrence of different words and on differences of meaning of words → what about their collocability?
 - example: uses of “Mobilität” (“unemployment” vs. “mobility”) (cf. Abel 2018)

<[...] erläuterten diese Initiative des Landes zugunsten der Arbeitnehmerinnen . Sie waren im Juli 2000 zusammen mit anderen **in die Mobilität entlassen** worden - die einen beim Naturnser Industriebetrieb Ivoclar , die anderen beim Meraner Textilunternehmen Merlet. >
(Dolomitenkorpus, 2001)

*(In June 2000, together with others, they were *released into mobility [literal translation], some at the industrial company Ivoclar in Naturns, some at the textile company Merlet in Meran.)*

Challenges

Examples:

- language change:
 - focus on neologisms and variants not yet lexicalised
 - example 1: meaning not yet lexicalised;
 - “Skonto” (“discount” – “discount when immediate cash payment”)

<„Nicht jede Apotheke gewährt den gleich großen **Skonto**“, erklärt Cimatti. „Das hängt von der Größe der Apotheke ab.“> (Dolomitenkorpus, 2001)

*(“Not every pharmacy allows the same **discount**”, explained Cimatti. “It depends on the size of the pharmacy.”)*

Challenges

Examples:

- language change:
 - focus on neologisms and variants not yet lexicalised
 - example 2: word form & meaning not yet lexicalised;
“Vollkornpizzetta” (“very small pizza made of whole grain”)



Challenges

Examples:

- language change:
 - focus on neologisms and variants not yet lexicalised
 - example 3: neologism (candidate);
word field around “Autonomie” (“autonomy” - in a political sense)

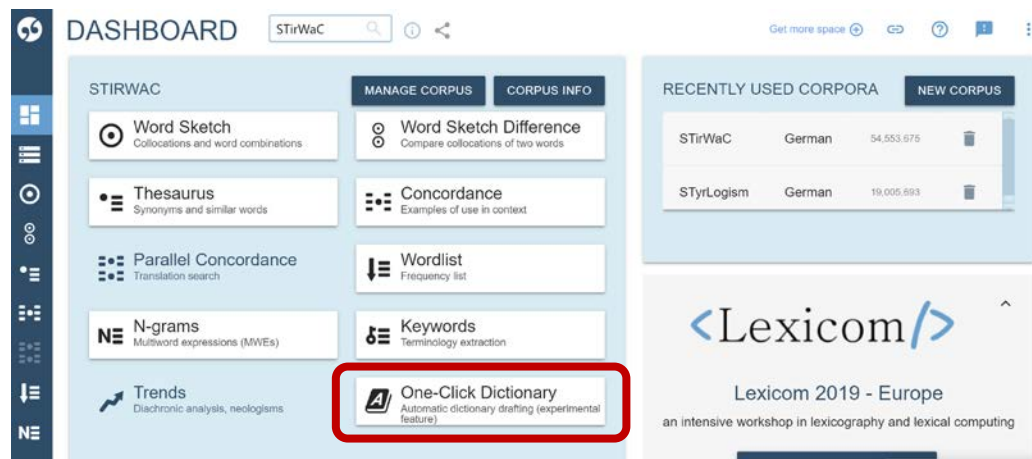
<Die Führung der SVP soll das Projekt einer **Vollautonomie** wieder intensiv vorantreiben >
(Dolomitenkorpus, 2001)

*(The leader of the SVP should put forward the project of a **full autonomy** more intensively again).*

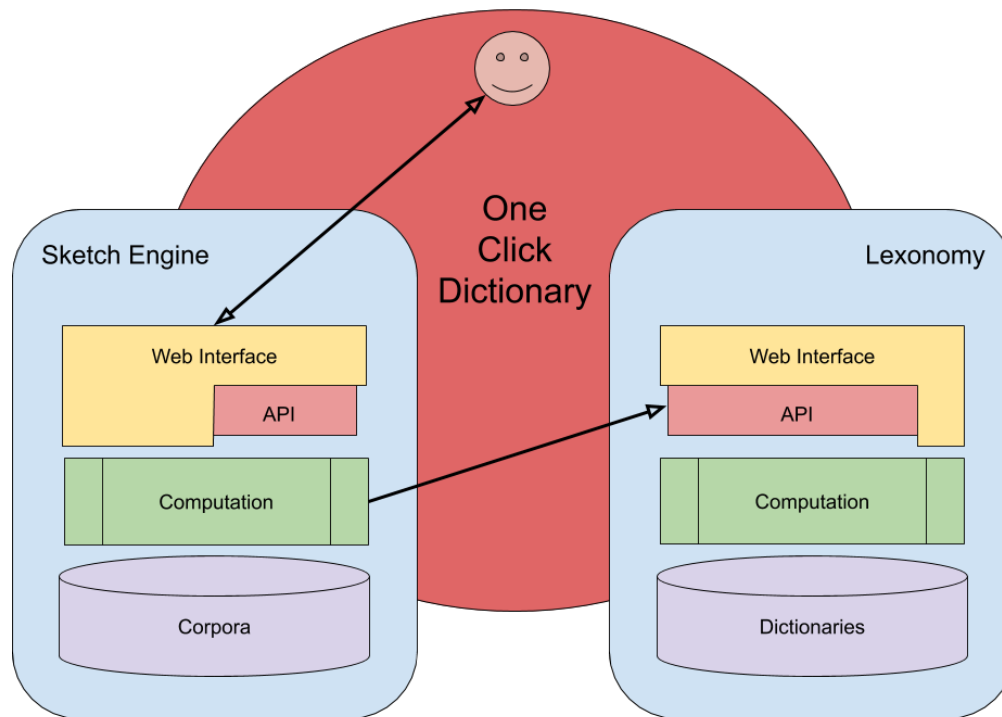
(Semi)-automated opportunities

Examples:

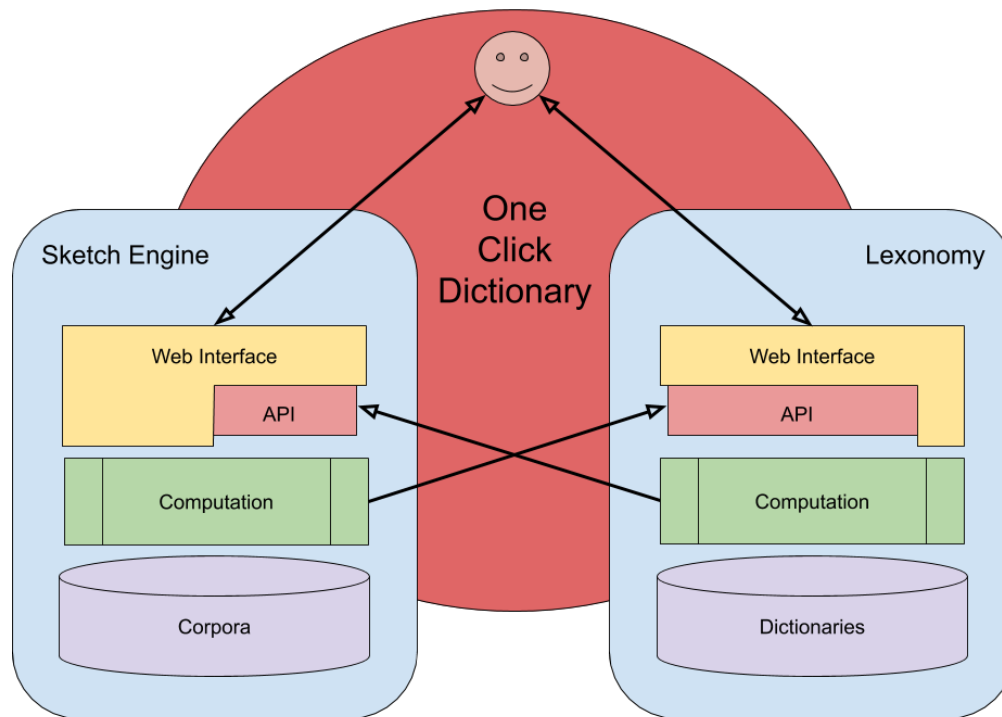
- lexicographic treatment of the standard German used in South Tyrol:
 - focus on the pipeline of producing a dictionary
 - example: “One-Click Dictionary” paradigm (Sketch Engine & Lexonomy) (<https://app.sketchengine.eu/>, <https://www.lexonomy.eu/>); illustrative showcase “Bar” (“coffee shop”)



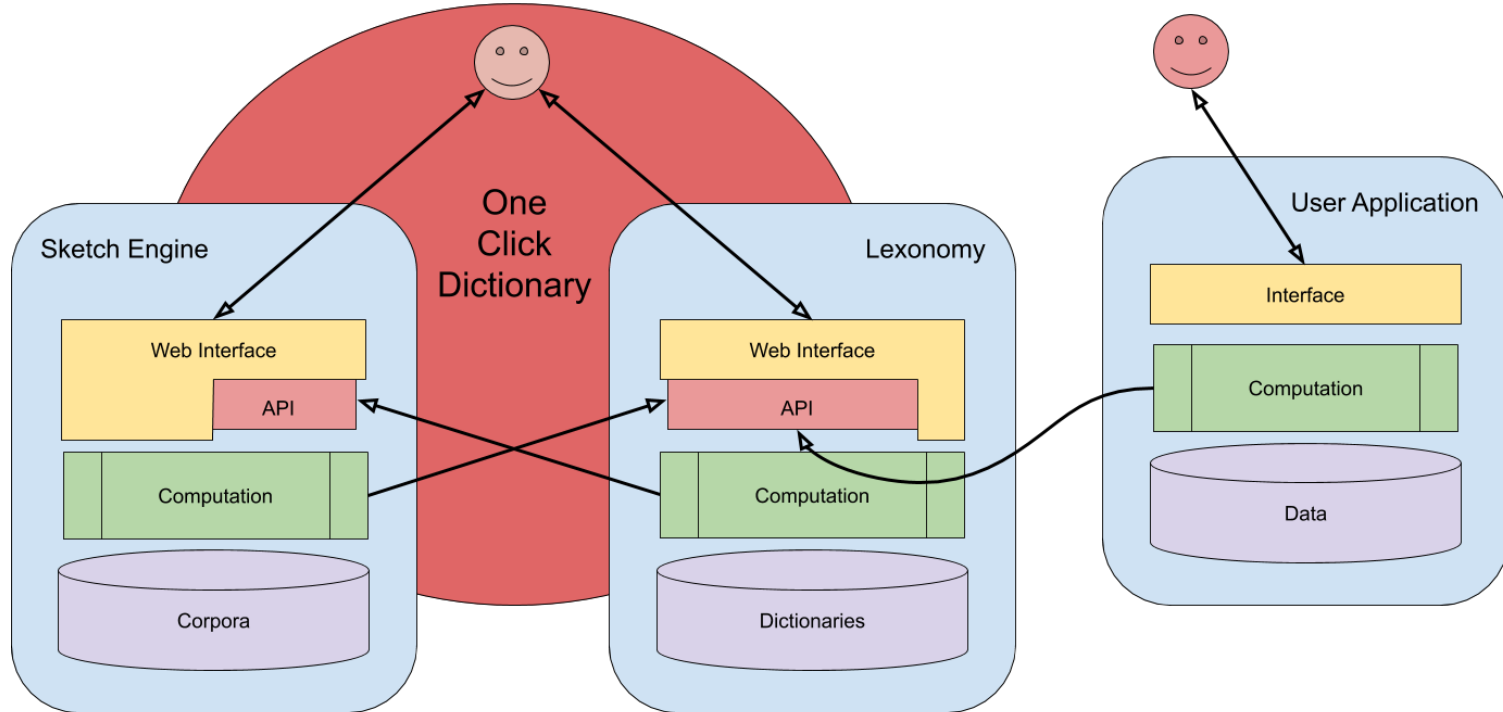
(Semi)-automated opportunities



(Semi)-automated opportunities



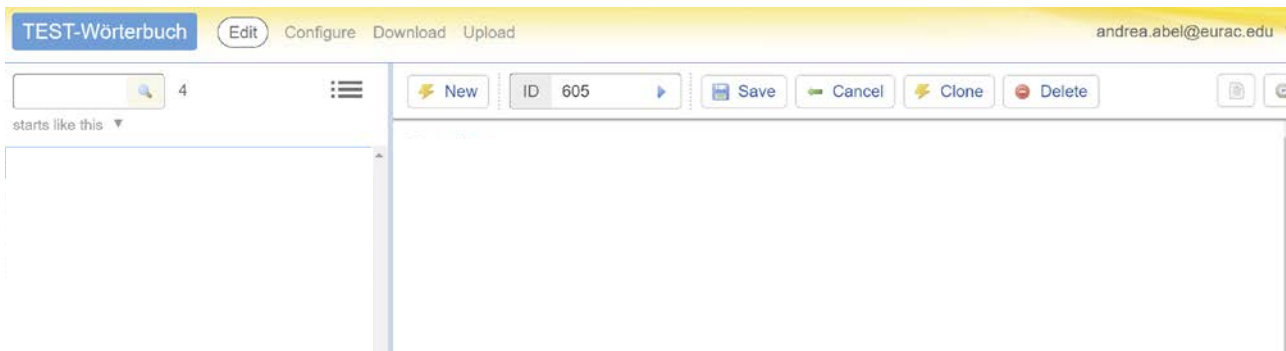
(Semi)-automated opportunities



(Semi)-automated opportunities

Examples:

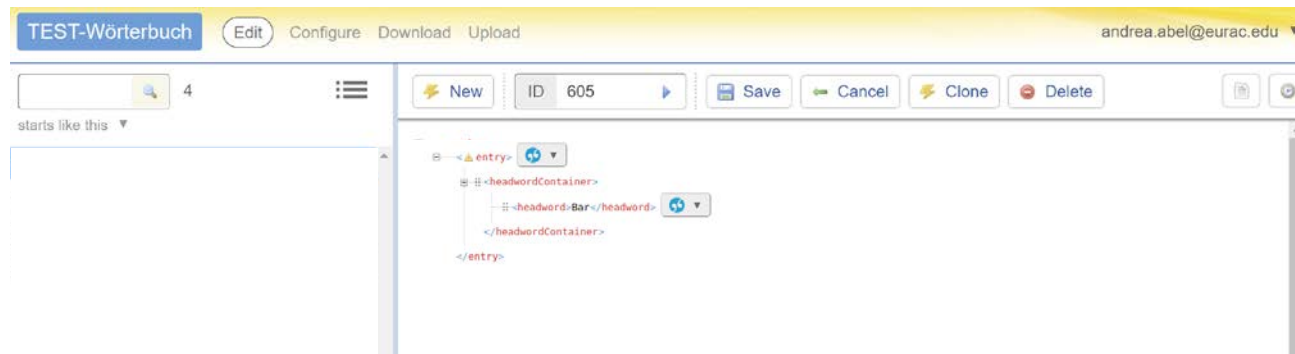
- lexicographic treatment of the standard German used in South Tyrol:
 - focus on the pipeline of producing a dictionary
 - example: “One-Click Dictionary” paradigm (Sketch Engine & Lexonomy) (<https://app.sketchengine.eu/>, <https://www.lexonomy.eu/>); illustrative showcase “Bar” (“coffee shop”)



(Semi)-automated opportunities

Examples:

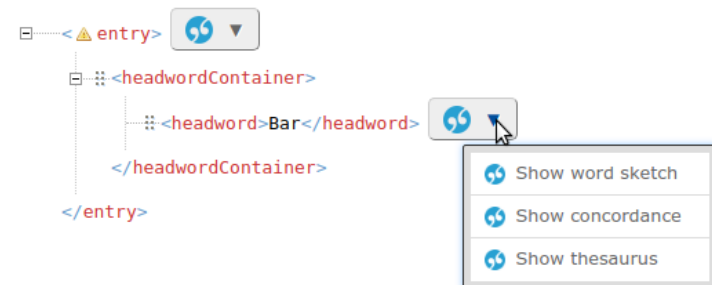
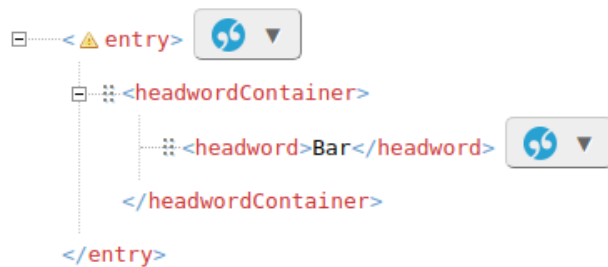
- lexicographic treatment of the standard German used in South Tyrol:
 - focus on the pipeline of producing a dictionary
 - example: “One-Click Dictionary” paradigm (Sketch Engine & Lexonomy) (<https://app.sketchengine.eu/>, <https://www.lexonomy.eu/>); illustrative showcase “Bar” (“coffee shop”)



(Semi)-automated opportunities

Examples:

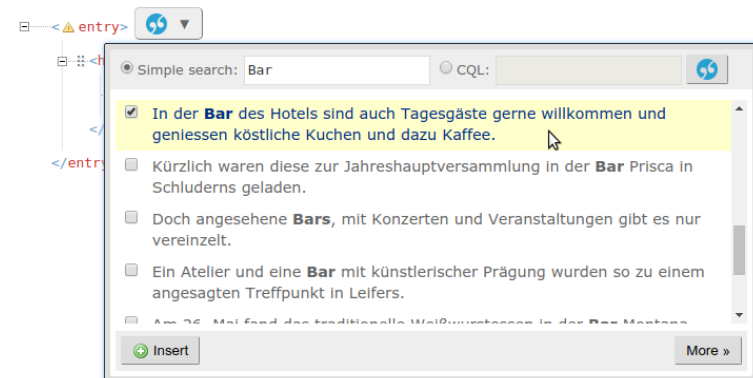
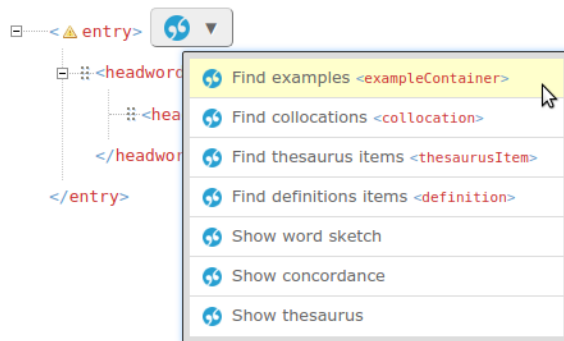
- lexicographic treatment of the standard German used in South Tyrol:
 - focus on the pipeline of producing a dictionary
 - example: “One-Click Dictionary” paradigm (Sketch Engine & Lexonomy) (<https://app.sketchengine.eu/>, <https://www.lexonomy.eu/>); illustrative showcase “Bar” (“coffee shop”)



(Semi)-automated opportunities

Examples:

- lexicographic treatment of the standard German used in South Tyrol:
 - focus on the pipeline of producing a dictionary
 - example: “One-Click Dictionary” paradigm (Sketch Engine & Lexonomy) (<https://app.sketchengine.eu/>, <https://www.lexonomy.eu/>); illustrative showcase “Bar” (“coffee shop”)



Data basis: STyrLogism corpus (webcorpus)

(Semi)-automated opportunities

Examples:

- lexicographic treatment of the standard German used in South Tyrol:
 - focus on the pipeline of producing a dictionary
 - example: “One-Click Dictionary” paradigm (Sketch Engine & Lexonomy) (<https://app.sketchengine.eu/>, <https://www.lexonomy.eu/>); illustrative showcase “Bar” (“coffee shop”)

The screenshot shows a web browser window with the URL <https://www.lexonomy.eu/ze4y3ccm8/edit/entry/>. The page title is "STyrLogism" and the user is logged in as "egon.stemle@eurac.edu". The interface includes a search bar with the number "3" and a list of entries starting with "Bar". The selected entry is "Bar", which is defined as "das Café | die Kaffeestube" and includes the example sentence: "In der Bar des Hotels sind auch Tagesgäste gerne willkommen und geniessen köstliche Kuchen und dazu Kaffee." The interface also features buttons for "New", "Edit", "Clone", and "Delete", along with an "ID" field showing "615".

Conclusions & outlook

- Needs
 - need of appropriate corpora to observe language use (including everyday situations) and detect trends of the local standard variety of the German language
 - need of support for automatically extracting relevant data for variety lexicography (e. g. collocations, “new” word forms and meanings)
 - need of easy-to-use tools

Conclusions & outlook

- Desirable developments through cooperations with an international lexicographic infrastructure (ELEXIS)
 - strengthen the position of local varieties and dialects
 - access to current methods and tools, awareness of challenges and upcoming solutions for the next generation of online dictionaries
 - possibility to integrate local digital resources into modern workflows and to test them together
 - influence on the design for tools and workflows

Thank you for your attention!



Andrea Abel
(andrea.abel@eurac.edu)

&

Egon Stemle
(egon.stemle@eurac.edu)



Institute for Applied Linguistics
(www.eurac.edu/iscm)