



universität
wien

Emmerich Kelih
Tatjana Vučajnk

Institut für Slawistik

Minimalwortschatz Slowenisch – Deutsch

Von der Theorie zu den praktischen Problemen der Erstellung

- Institut für Slawistik, Universität Wien
- Institut für Slawistik, Universität Graz
- <https://slawistik.univie.ac.at/mitarbeiter/kelih-emmerich//> [emmerich.kelih@univie.ac.at]

Überblick

1. Einleitung
2. Grund- und Aufbauwortschatz: Projektskizze
3. Auswahl der Themenbereiche
4. Prinzipien der Annotation
5. grammatische Annotation
6. phonetische Auszeichnung (Wortakzent)
7. Rolle der Vorkommenshäufigkeit
8. Häufigkeitswörterbücher: Slowenisch
9. Ausblick und Zusammenfassung

Einleitung:

- gestiegenes Interesse an Lehrbüchern Slowenisch als L2
(insbesondere in Kärnten und der Steiermark)
- positive Entwicklung: sowohl im schulischen als auch im universitären Bereich, Erwachsenenbildung
- Entwicklung von Lehrmaterialien (Center za slovenščino kot drugi/tuji jezik – Center for Slovene as a second/foreign language)
- u.a.: Ferbežar, I. et al (2004): *Sporazumevalni prag za slovenščino*. Ljubljana: Center za slovenščino kot drugi/tuji jezik. [= Threshold Level for Slovenian]

→ Grund- und Aufbauwortschatz Slowenisch – Deutsch

→ basierend auf **slowenischer Standardsprache**, aber unter Berücksichtigung der gesprochenen Umgangssprache

Grund- und Aufbauwortschatz: Projektskizze

Grundwortschatz: ca. **3000** Lemmata

Aufbauwortschatz: ca. **2000** Lemmata

Lemma (slowenisch, annotiert)

Übersetzung (deutsch)

Beispielsatz (slowenisch)

Beispielsatz (deutsch)

GERS (Gemeinsame europäische Referenzrahmen für Sprachen): A1 – B2, z.T. bis C1

Themenbereiche nach:

Glaboniat, M.; Müller, M. Rusch P. (2002): *Profile Deutsch*. Berlin: Langenscheidt.

Themenbereiche:

- (1) Personalien, Informationen zur Person
- (2) Wohnen
- (3) Umwelt, Pflanzen, Tiere
- (4) Reisen und Verkehr
- (5) Speisen und Verpflegung
- (6) Einkaufen
- (7) Öffentliche und private Dienstleistungen
- (8) Körper, Gesundheit und Hygiene
- (9) Wahrnehmung und Motorik
- (10) Ausbildung und Schule
- (11) Arbeit und Beruf
- (12) Sprachen und Fremdsprachen
- (13) Freizeit und Unterhaltung
- (14) Persönliche Beziehungen und Kontakte
- (15) Politik und Gesellschaft
- (16) allgemeine Begriffe
- (17) häufigste Verben
- (18) Strukturwörter

| INHALT | |
|------------------------------------------|-----------|
| Zum neuen Grundwortschatz Englisch | 9 |
| 1 THEMENBEZOGENE BEGRIFFE | 13 |
| 1.1 Der Mensch | 13 |
| 1.1.1 Körper und Wesen | 13 |
| 1.1.1.1 Körper | 13 |
| 1.1.1.2 Aussehen | 17 |
| 1.1.1.3 Geist und Verstand | 19 |
| 1.1.1.4 Charakter | 23 |
| 1.1.1.5 Fühlen und Empfinden | 26 |
| 1.1.1.6 Gesundheit und Krankheit | 33 |
| 1.1.1.7 Leben und Tod | 37 |

| | |
|------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| tooth [tu:θ] <i>n</i> <i>pl</i> teeth [ti:θ] It's important to brush your teeth after every meal. | Zahn <i>m</i> Es ist wichtig, sich nach jeder Mahlzeit die Zähne zu putzen. |
| tongue [tʌŋ] <i>n</i> The soup was so hot that I burned my tongue . | Zunge <i>f</i> Die Suppe war so heiß, dass ich mir die Zunge verbrannte. |
| neck [nek] <i>n</i> He put his arms round her neck and kissed her. | Hals <i>m</i> , Nacken <i>m</i> , Genick <i>n</i> Er legte die Arme um ihren Hals und küsste sie. |
| chest [tʃest] <i>n</i> He had sharp pains in the chest . | Brust <i>f</i> , Brustkorb <i>m</i> Er hatte heftige Schmerzen im Brustkorb. |

aus:

Langenscheidt Grundwortschatz Englisch
(2000)

→ erste Orientierung für Auswahl und Strukturierung!

Grammatikalische Annotation der Lemma

Nomen, Verben, Adjektive, Pronomen, Partikeln, Interjektionen

Wortarten und Form der Lemmata nach:

- Slovar slovenskega knjižnega jezika (1970) und Slovenski pravopis (2001)
- Deskriptoren und Qualifikatoren werden als Abkürzungen eingeführt
- Ziel: wichtigste grammatikalische Information zu den Lemma zu liefern
- Berücksichtigung von Varianten nur im geringeren Ausmaß (wie z.B. **grád** -ú in -a)
- deutsche Übersetzung orientiert sich an PONS Wörterbuch (Slowenisch-Deutsch, Deutsch-Slowenisch), 2008: Verlag: Rokus-Klett.

Grammatikalische Annotation der Lemma

Nomen

sín -a *m pl* sinóvi

Sohn m.

Ôče in sín se ne razúmeta.

Vater und Sohn verstehen sich nicht.

Verb

izglédati -am *impf*

aussehen

Ôna dôbro izgléda. Dôbro izglédaš! - Sie sieht gut aus. Gut siehst du aus!

* biaspektuellen Verben, wie z.B. - rodíti *impf / pf* (gebären) werden als solche gekennzeichnet

Polysemie

- keine besondere Kennzeichnung

- Auswahl einer bestimmten Bedeutung erfolgt z.T. intuitiv (lastni jezikovni čut)
- wichtige bzw. häufig empfundene „Bedeutung“

poznáti -znám se *impf*

(1) im Sinne „von jemanden kennen“

(2) „um die Existenz, Eigenschaft, Besonderheit von etwas Bescheid wissen“

Ali se poznáta? - Kennt ihr beiden euch?

Kollokationen

Wortverbindungen, die in der Regel gemeinsam auftreten („stalne besedne zveze“)

- **hišna** (poštna, telefonska) **številka**
- **rojstni dan, dekliški priimek**
- **elektronski naslov, stalno prebivališče**

* kommunikative Kompetenz (Ämter, Behörden, Ausfüllen von Formularen)

- **živeti skupaj**

als umgangssprachliche Form für den Ausdruck von „in einer Lebensgemeinschaft sich befinden“

* kommunikative Kompetenz: Umgangssprache

Akzentuierung

- freie, dynamische Betonung (jede einzelne Betonungsposition muss gesondert erwerben)

Annotation: exspiratorische Akzent (keine Toneme!)

Akut (ostrivec) - **stárost** tudi **staróst** (alternative Akzentuierung möglich)

Stárost je lahkó lépa. - Das Alter kann schön sein.

Brevis (krativec) - **kràj** krája (betontes langes a im Gen. Pl.)

V katérem kráju živiš? – In welchem Ort lebst du?

Zirkumflex (strešica) - **jêzik** jezíka (Betonungsposition, Vokalqualität)

Katére jezíke govoriš? - Welche Sprachen sprichst du?

→ Hohe Akzentvariabilität im Slowenischen!

Das Slowenische: Zwischen Norm und Usus

Standardsprache („zborni jezik“): Standardsprache in der gesprochenen Form („splošnopogovorni jezik“)

Praxis- und Lebensbezug der angeführten Beispielsätze

z.B. Gastronomie

- **natakar** – gostinski delavec, ki se ukvarja s strežbo gostov
- **kelnar** umgs. (nekj. ljud. – SP; nižje. pog. SSKJ) *natakar*
- **točaj** (SSKJ, SP) - gostinski delavec, ki streže gostom pijačo
- **šank** nižje pog. *točilna miza* (SSKJ), neknoj. pog. *točilni pult* (SP)

lušten pog. čeden, ljubek (SSKJ), nekj. pog. (SP) - hübsch, staatlich

Imèl je lúštno púnco. – Er hatte eine hübsche Freundin.

Häufig vorkommende umgangssprachliche Ausdrücke

zavóra - **brémza** *umgs.*

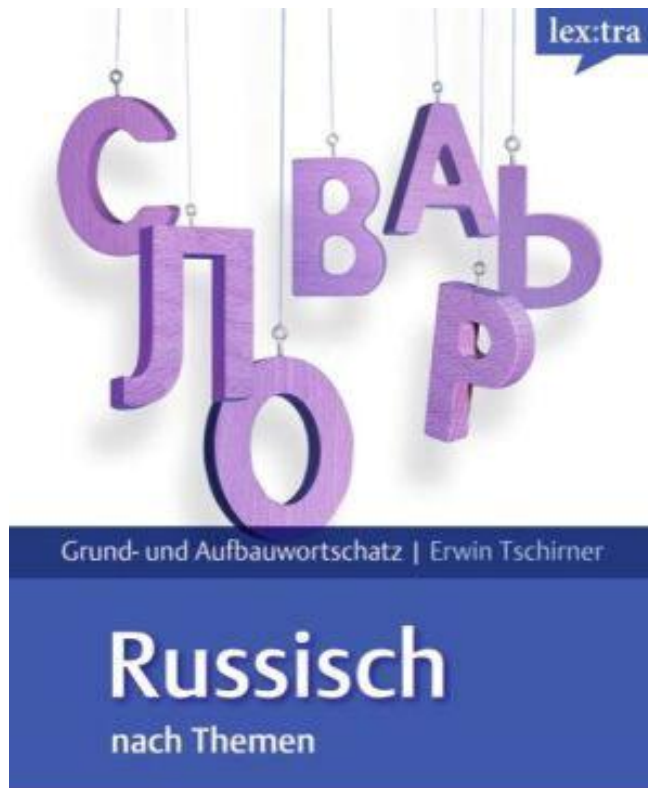
zavírati - **brémzati** *umgs.*

Ne brémzaj kar napréj!

Zavírati je knjížni izràz za brémzati. – “Zavirati“ ist der standardsprachliche Ausdruck für „zu bremsen“.

Auswahl der Lemmata: „Gewichtung“ durch Vorkommenshäufigkeit

Tschirner, Erwin (2010): *Grund- und Aufbauwortschatz. Russisch nach Themen*. Berlin: Cornelsen.



→ lexikalisches Minimum

→ thematisch gegliedert

→ ca. 2000 Lemmata

→ basiert auf:

Sharoff, S. (2002): The frequency dictionary for Russian →

<http://bokrcorpora.narod.ru/frqlist/frqlist-en.html>

Национальный корпус русского языка
/Nacional'nyj korpus russkogo jazyka

Häufigkeitswörterbücher für Slowenisch?

- weitaus schlechtere Ausgangslage als für das Russisch, das Tschechische, das Slowakisch usw.
- kein „repräsentatives“ Häufigkeitswörterbuch für das Slowenische
- z.T. Absichtserklärungen (T. Erjavec, P. Jakopin)

Jakopin, Primož (1995): Nekaj številik iz Slovarja Slovenskega Knjižnega Jezika. In: Slavistična revija 43 (3), S. 341–375.

Liste der 500 häufigsten Wortformen aus dem Slovar Slovenskega Knjižnega Jezika 1970 (SSKJ 1970)

Jakopin, Primož (2002): Entropija v slovenskih leposlovnih besedilih. Ljubljana: Založba ZRC.

→ Wortlisten für zwei Prosa-Korpora des Slowenischen

- im Korpus „Nova Beseda“ besteht die Möglichkeit der Abfrage nach der Häufigkeit von Wortformen
- aus den bestehenden Korpora (Fidaplust, Gigafida, Gos) kann nur die Frequenz einzelner Wortformen extrahiert werden

→ nicht befriedigende Situation !

Erstellung eines Häufigkeitswörterbüchern

1. Auswahl der Texte (Textsorten, Funktionalstile)
2. Extrahierung von Wortformen: Tokens - Types
3. Frage der Lemmatisierung (manuell vs. automatisiert)

Erstellung eines Häufigkeitswörterbuch: Slowenisch

- basiert auf unterschiedlichen Funktionalstilen/Diskurstypen
- ca. 1600 Texte
- keine älteren Sprachzustände berücksichtigt
- grobes Tagging

| Textsorte | Tokens | Types | TTR |
|------------------|---------------|--------------|------------|
| Administration | 49343 | 11722 | 4,21 |
| Alltag | 51327 | 8854 | 5,80 |
| Drama | 23935 | 5847 | 4,09 |
| Gedichte | 53377 | 13289 | 4,02 |
| Journalistik | 176198 | 33348 | 5,28 |
| Prosa | 881167 | 66614 | 13,23 |
| Wissenschaft | 179865 | 32063 | 5,61 |
| Gesamt | 1216374 | 109211 | 11,14 |

Häufigste Wortformen (Types)

| | | | | | | | | | |
|----|------|----|-------|----|------|----|--------|-----|-------|
| 1 | JE | 21 | ALI | 41 | BILA | 61 | MED | 81 | BOM |
| 2 | IN | 22 | TO | 42 | KOT | 62 | JI | 82 | PROTI |
| 3 | SE | 23 | GA | 43 | O | 63 | TA | 83 | BOLJ |
| 4 | V | 24 | KO | 44 | JO | 64 | A | 84 | TAM |
| 5 | DA | 25 | SI | 45 | KAR | 65 | TER | 85 | BILI |
| 6 | NA | 26 | BIL | 46 | LE | 66 | TEM | 86 | OČI |
| 7 | NE | 27 | KAKOR | 47 | ME | 67 | LAHKO | 87 | SVA |
| 8 | PA | 28 | OD | 48 | KER | 68 | PRAV | 88 | RES |
| 9 | SO | 29 | MU | 49 | JIH | 69 | SMO | 89 | TISOČ |
| 10 | ZA | 30 | ŽE | 50 | KAKO | 70 | NEKAJ | 90 | DOBRO |
| 11 | BI | 31 | BILO | 51 | VEČ | 71 | NIČ | 91 | SVOJO |
| 12 | KI | 32 | KAJ | 52 | PRED | 72 | ZATO | 92 | DAN |
| 13 | NI | 33 | BO | 53 | NAJ | 73 | POTEM | 93 | BOG |
| 14 | Z | 34 | MI | 54 | TE | 74 | SAJ | 94 | NISO |
| 15 | SEM | 35 | IZ | 55 | STA | 75 | SAM | 95 | NISEM |
| 16 | PO | 36 | TI | 56 | JAZ | 76 | TEGA | 96 | VSI |
| 17 | ŠE | 37 | PRI | 57 | SAMO | 77 | BREZ | 97 | KJER |
| 18 | TUDI | 38 | ČE | 58 | OB | 78 | POD | 98 | REKEL |
| 19 | TAKO | 39 | VSE | 59 | ZDAJ | 79 | SVOJE | 99 | NAS |
| 20 | S | 40 | DO | 60 | K | 80 | VENDAR | 100 | NAM |

Nächstes Ziel: Lemmatisierung

Ausblick und Zusammenfassung

- 5 Kapitel sind bereits fertig
- z.T. ist die Zuordnung von Lemmata zu Grund- und Aufbauwortschatz noch nicht abgeschlossen
- Beispielsätze und die Auswahl ist noch zu diskutieren
- Registererstellung
- leserfreundliche Formatierung
- Projektfertigstellung: Anfang 2014

Sponsoren des Projektes



Layout, Lektorat, Vertrieb

Literatur in Auswahl:

Freese, Holger; et al. (2000): Grundwortschatz Englisch. Ein nach Sachgebieten geordnetes Lernwörterbuch mit Satzbeispielen. Berlin u.a.: Langenscheidt.

Jakopin, Primož (2002): Entropija v slovenskih leposlovnih besedilih. Ljubljana: Založba ZRC.

Jakopin, Primož (2006): On Text Corpora, Word Lengths, and Word Frequencies in Slovenian. In: Peter Grzybek (Hg.): Contributions to the Science of Language. Word Length Studies and Related Issues. Boston u.a.: Kluwer, S. 171–186.

Glaboniat, M.; Müller, M. Rusch P. (2002): Profile Deutsch. Berlin: Langenscheidt.

Hamann, Carola; Wienecke, Natalia (2008): Grundwortschatz Russisch. München: Hueber.

Slovar Slovenskega Knjižnega jezika (1970ff). Ljubljana.

Slovenski Pravopis (2001). Ljubljana: SAZU.

Tschirner, Erwin (2010): Grund- und Aufbauwortschatz. Russisch nach Themen. Berlin: Cornelsen.