

# **Retrodigitalisierung von Wörterbüchern mit diskursiver Mikrostruktur**

## **Lexikografische und technische Aspekte am Beispiel eines Lehnwörterbuchs**

Peter Meyer

Institut für Deutsche Sprache

meyer@ids-mannheim.de

# Fahrplan

- Grundsätzliche Informationen zum IDS-Projekt „Lehnwortportal“
- technische Umsetzung des Internetportals
- Aufbereitung eines älteren Print-Lehnwörterbuchs mit diskursiver Mikrostruktur
- lexikografische Probleme und Fazit

## IDS-Projekt „Lehnwortportal“

### Eckdaten

- Projekt „Einrichtung eines Internetportals für deutsche Lehnwörter in slawischen Sprachen“
- gefördert durch den Beauftragten der Bundesregierung für Kultur und Medien
- Pilotprojekt für geplantes umfangreicheres Portal

### Grundidee

- Portal von Lehnwörterbüchern **mit Deutsch als gemeinsamer Gebersprache**
- alle Wörterbücher einzeln nutzbar
- zusätzlich wörterbuchübergreifende Suchmöglichkeiten...
  - insbesondere ausgehend von deutschen Etyma – „**umgekehrtes Lehnwörterbuch**“ (Engelberg 2010)

# Grundkonzeption des Portals

## Wörterbücher des Pilotprojekts

Vincenz, A.; Hentschel, G. (2010): Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache.

Menzel, T.; Hentschel, G. (2005). Wörterbuch der deutschen Lehnwörter im Teschener Dialekt des Polnischen. Oldenburg.

Striedter-Temps, Hildegard (1963): Deutsche Lehnwörter im Slovenischen. Wiesbaden: Harrassowitz.

## weitere Wörterbücher (geplant)

Engelberg, S. et al. : An inverted loanword dictionary of German loanwords in the languages of the South Pacific.

[Wb deutscher Lehnwörter im Weißrussischen]

[Japanisch, Schwedisch, Usbekisch, Russisch, ...]

## Wörterbücher als Stand-alone-Ressourcen

## Deutsche Metalemmaliste

A...

B...

C...

D...

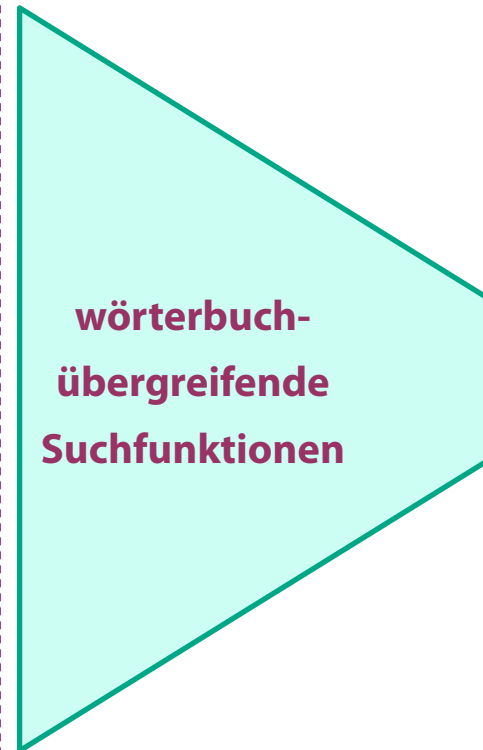
E...

F...

G...

H...

...



# Nutzungsmöglichkeiten

## Nutzung des Internetportals

- **alphabetische Lemmalisten:** Lemmata der Einzelwörterbücher; wörterbuchübergreifende Metalemmaliste der deutschen Etyma
- **Suche über Webseitenformulare:** nach einzelnen Lehnwörtern / Etyma inklusive Teilzeichenketten; nach Zeitraum / Sprache / geographischem Raum / grammatischen Merkmalen / ...
  - *Ausbau:* Suche nach dem Entlehnungsweg; nach semantischen Kennzeichen

## Wissenschaftliche Nutzungsmöglichkeiten der Portaldaten

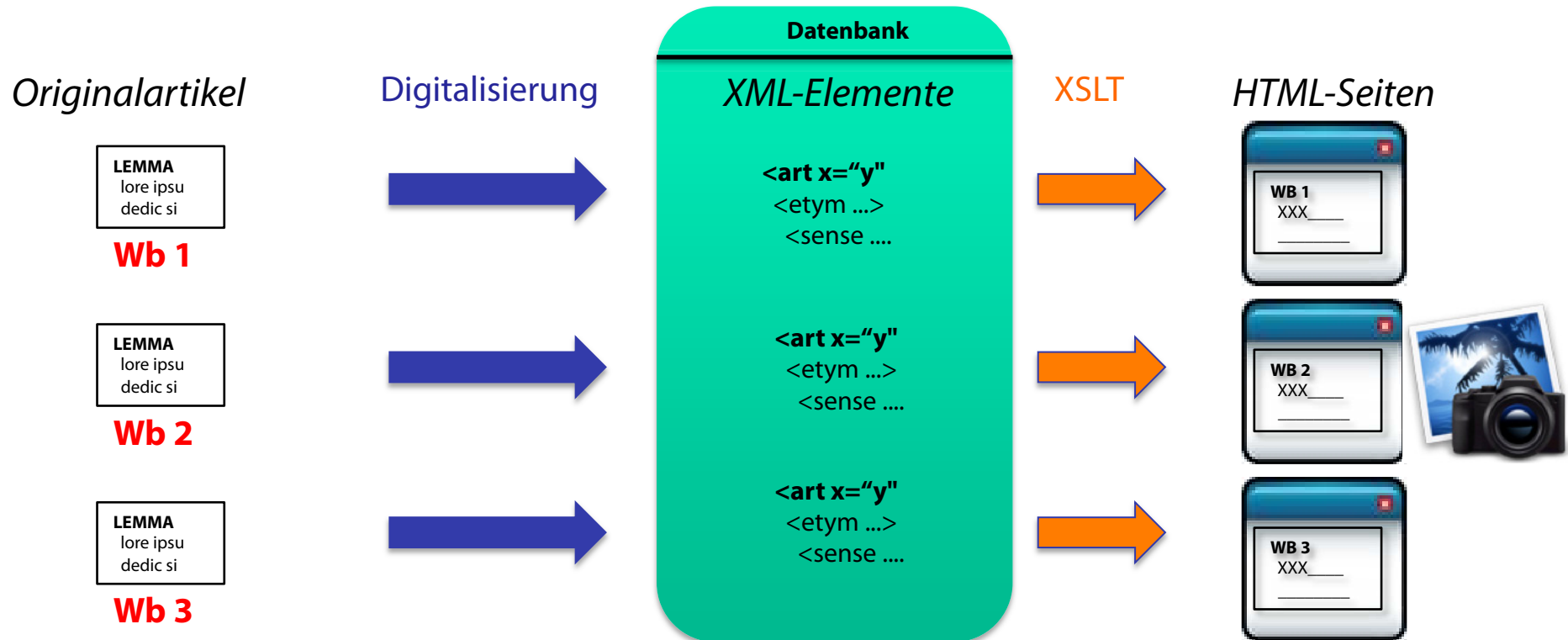
- **Eigendynamik von Entlehnungsprozessen:** sprachbezogener Zeitverlauf der Entlehnungsfrequenzen von Lexemen
- **Sprachkontakttypbedingte zeitliche und ontologische Lehnwortverteilungen:** Entlehnungszeitverläufe und spezifische Muster der Lehnwortstreuung zur Unterscheidung von Sprachkontakttypen
- **Lexikalische Germanoversalien:** Machen bestimmte phonologische, morphologische oder semantische Eigenschaften bestimmte Wörter aus dem Deutschen besonders entlehnungsfreundlich?
- **Lehnwortpersistenz:** Welche sprachlichen und soziopolitischen Faktoren führen zur Stabilisierung bzw. Destabilisierung von Lehnwortgut im Lexikon?

---

# Fahrplan

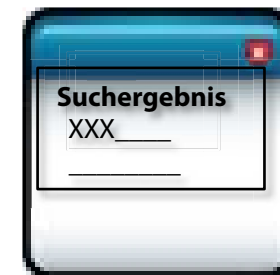
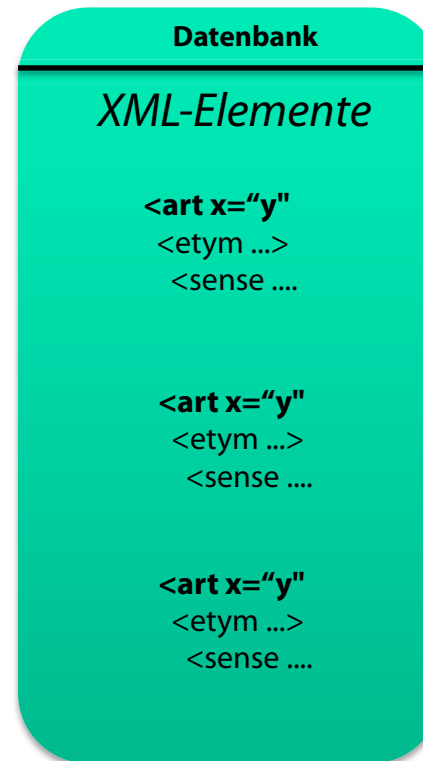
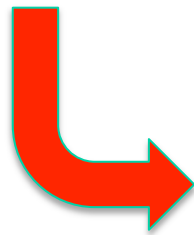
- Grundsätzliche Informationen zum IDS-Projekt „Lehnwortportal“
  - technische Umsetzung des Internetportals
  - Aufbereitung eines älteren Print-Lehnwörterbuchs mit diskursiver Mikrostruktur
  - lexikografische Probleme und Fazit
-

# XML-basierte Repräsentation der Einzelwörterbücher



# XML-basierte Repräsentation der Einzelwörterbücher

**Abfrage:** „Finde alle polnischen und slovenischen maskulinen Substantive zu deutschen Etyma auf *-ung*, die vor 1500 entlehnt wurden“



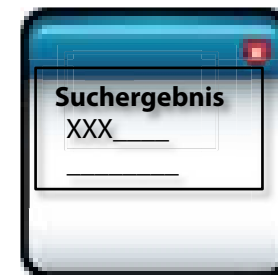
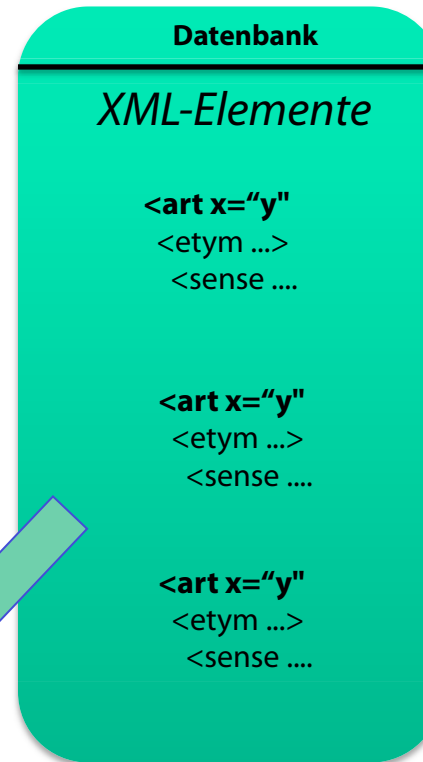
## Probleme mit direkten wb-übergreifenden Abfragen über XML-Dokumenten :

- „teuer“: lange Antwortzeiten und hohe Auslastung des Datenbanksystems
- inkompatible XML-Schemata, Angabeformate in verschiedenen Wörterbüchern
- Metalemmalisten deutscher Etyma umfangreiche lexikografische Aufgabe
- ...



# XML-basierte Repräsentation der Einzelwörterbücher

**Abfrage:** „Finde alle polnischen und slovenischen maskulinen Substantive zu deutschen Etyma auf *-ung*, die vor 1500 entlehnt wurden“



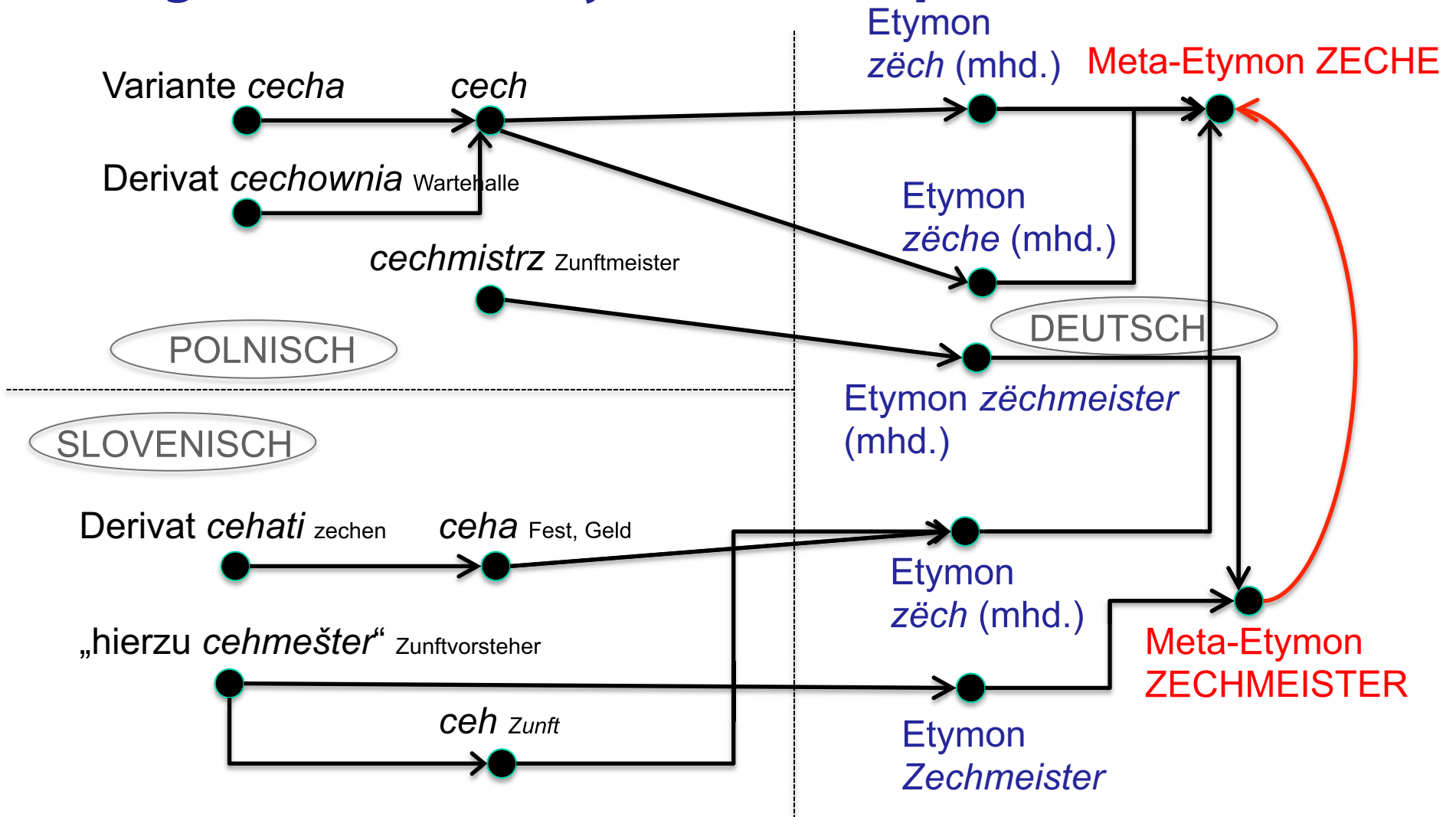
**Lösung:** zusätzliche relationale Datenbanktabellen für wb-übergreifenden Zugriff auf ausgewählte lexikografische Daten sowie auf deutsche Metalemmaliste



# Notwendigkeit wörterbuchübergreifender lexikographischer Datenverwaltung

- **Universelles XML-Schema für alle erdenklichen Wörterbuchformate illusorisch**
  - realistisch sind nur wenige allgemeine Anforderungen an Mikro-/Makrostruktur
  - ...darunter Konvertierbarkeit in TEI.dictionaries-Format
- **unterschiedliche Klassifikationen/Ontologien in den Einzelwörterbüchern**
  - diasystematische Einordnung (Dialektbezeichnungen, Sprachstufen)
  - Granularität zeitlicher Bestimmungen (Entlehnungszeitraum, Erstbelegdatum)
  - Morphosyntax: grammatische Kategorien
- **Metalemmaliste für Etyma erforderlich**
  - Identität von dt. Etyma innerhalb von und zwischen den Portal-Wörterbüchern (polnisch *cech* ‚Gilde, Geldbeitrag‘, slovenisch *ceh* ‚Zunft‘, sloven. *ceha* ‚Gelage, Zechgeld‘, dt. *Zeche*)?
  - Verwaltung von ‚Mittelgliedern‘ in Entlehnungsketten – Identität der Etyma
  - vom dt. Etymon *Jammer* aus auch poln. *kacenjamer* auffindbar?

# Lösung: Gerichteter azyklischer Graph (DAG)



● = Portalinstanz: Wort(form) im Artikel mit portalweit standardisierten Informationen

## Erstellung der Metalemmaliste deutscher Etyma



- knapp 5000 Artikel im Rahmen dieses Pilotprojekts
- Bearbeitung der Metaetymaliste schon bei diesem Umfang nur softwaregestützt möglich



---

# Fahrplan

- Grundsätzliche Informationen zum IDS-Projekt „Lehnwortportal“
  - technische Umsetzung des Internetportals
  - **Aufbereitung eines älteren Print-Lehnwörterbuchs mit diskursiver Mikrostruktur**
  - lexikografische Probleme und Fazit
-

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, *-am se* vb.impf., *traga se mi* = *utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, *-íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati*, *-ga* vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW.360.

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; **davon:** *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, *-am se* vb.impf., *traga se mi* = *utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, *-íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati*, *-ga* vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW.360.

→ Identifikation von Angabetypen schwierig

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *traqljivic* »Faulenzer« und *traqljívost* »Faulheit« (Murko II, 676); *trágati se, -am se vb. impf.*, *traga se mi = utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív, -íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati, -ga vb. impf.*, *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II, 680, 740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen, trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I, 239; II, 1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW. 360.

→ grammatische Angaben fehlen häufig

→ direkte Entlehnung oder innerslovenische Ableitung?



## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, -am se vb.impf., *traga se mi = utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, -íva »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrâgati*, -ga vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW.360.

→ unnötige Ambiguitäten; Interpretations- und Recherchebedarf

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, *-am se* vb.impf., *traga se mi* = *utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, *-íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati*, *-ga* vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); **die angeführten Wörter gehen alle auf dt. *träge* zurück:** mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW.360.

→ irreführende Erläuterungen

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, *-am se* vb.impf., *traga se mi* = *utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, *-íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati*, *-ga* vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); die angeführten Wörter gehen alle auf dt. *träge* zurück: **mhd. *træge* adj.** »träge, langsam, verdrossen«, ***trâge* adv.** »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW.360.

→ Konsultation des lautgeschichtlichen Teils erforderlich

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, *-am se* vb.impf., *traga se mi* = *utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, *-íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati*, *-ga* vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); **ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*.** Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW.360.

→ Entlehnungszeitraum unklar

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II,676); *trágati se*, *-am se* vb.impf., *traga se mi* = *utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív*, *-íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati*, *-ga* vb.impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II,680,740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen*, *trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I,239; II,1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. **Die slov. Formen zeigen Suffix-Erweiterung.** Vgl. auch Mikl. EW.360.

→ irreführende Erläuterungen

## Striedter-Temps (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

*trágljiv* adj. »träge, faul«; davon: *tragljivic* »Faulenzer« und *tragljívost* »Faulheit« (Murko II, 676) *trágati se, -am se* vb. impf., *traga se mi = utrâga me* »es beliebt mir nicht«, Tolm., BlKr.; vgl. *utragljív, -íva* »träge«, ung.slov., östl. u. südöstl. Steierm.; *utragljívost* f. »Trägheit«, östl. u. südöstl. Steierm.; *utrágati, -ga* vb. impf., *utrâga me* »es beliebt mir nicht«, südöstl. Steierm.; *utrâga se mi*, Innerkr., Savinska dol. (Plet. II, 680, 740); die angeführten Wörter gehen alle auf dt. *träge* zurück: mhd. *træge* adj. »träge, langsam, verdrossen«, *trâge* adv. »mit Trägheit, Langsamkeit«; *trâgen, trægen* »träge sein oder werden«, unpers. mit Dat. »verdrießen«; *betrâgen*, unpers. »langweilen, verdrießen, nicht gelüsten«, *mih betrâget* »ich habe keine Lust« (Lex. I, 239; II, 1491); ahd. *trâgi*, adj., *trâgo* adv.; nhd. *träge*. Die slov. Formen zeigen Suffix-Erweiterung. Vgl. auch Mikl. EW. 360.

→ meist direkte Übernahme von Angaben, Transkriptionen, orthografischen Gepflogenheiten aus den verschiedenen Quellen

---

# Integration von Striedter-Temps (1963) in das Portal

## Probleme

- automatisches Parsing nicht möglich; auch keine automatisierte Erstellung eines strukturierten Rohdigitalisats möglich
- Einzelheiten:
  - **diskursiver Artikelaufbau**, keine formalisierte Mikrostruktur
  - **Orthographie und phonetische Transkriptionen** werden aus den sehr heterogenen lexikografischen Quellen übernommen; schon OCR äußerst problematisch
  - **Angabetypen** nicht klar abgrenzbar und aufzählbar
  - **Skopusambiguitäten**; generell hoher Interpretationsbedarf
  - keine klaren **Lemmatisierungs**prinzipien

## Pragmatische Lösung

- Einzelartikel werden als **Bilddigitalisate** im Portal präsentiert
  - für wörterbuchspezifische und portalweite Suchvorgänge müssen zentrale lexikografische Daten dennoch in datenzentrierten XML-Instanzen abgelegt werden → partielle **manuelle Extraktion** unumgänglich
-

# Teilweise manuelle Retrodigitalisierung 1

**Laufende Nr. → 0782**

**Lemma oder Wortform → kajša**

## Slovenische Formen

**Wortart → n**

Formen/Varianten	Morphosyntax
kajša	f
kajža	f

## Bedeutungen

- Bauernhaus
- die Keische, Hütte

## Deutsche Etyma

Metalemma(?/*)	areal/zeitl.	Jahr	Formen	Wortart	Morphos.	Bedeutung	Plausib.
	mhd.		<ul style="list-style-type: none"> <li>• k{umlautdachte n}se</li> <li>• k{umlautdachte n}sse</li> </ul>	<ul style="list-style-type: none"> <li>• n</li> <li>• n</li> </ul>	<ul style="list-style-type: none"> <li>•</li> <li>•</li> </ul>		
	mhd.		• k{umlautdachte n}sche	• n	•		
	mhd.		• kîsche	• n	•		
	spmh.		• khai{zetpunktoben}e	• n	•	Gehäuse	

## Innerslovenische Ableitungen und Weiterbildungen zu den Formen

Form	areal/zeitl.	Wortart	Morphosyntax	Bedeutung	Plausib.
kaušlar		n		Bewohner, Besitzer einer Keische	

Jede Gruppe etymologisch einheitlich bewerteter Formen eines Artikels wird jeweils auf einem **Erhebungsbogen** (Word) registriert



## Teilweise manuelle Retrodigitalisierung 2

```

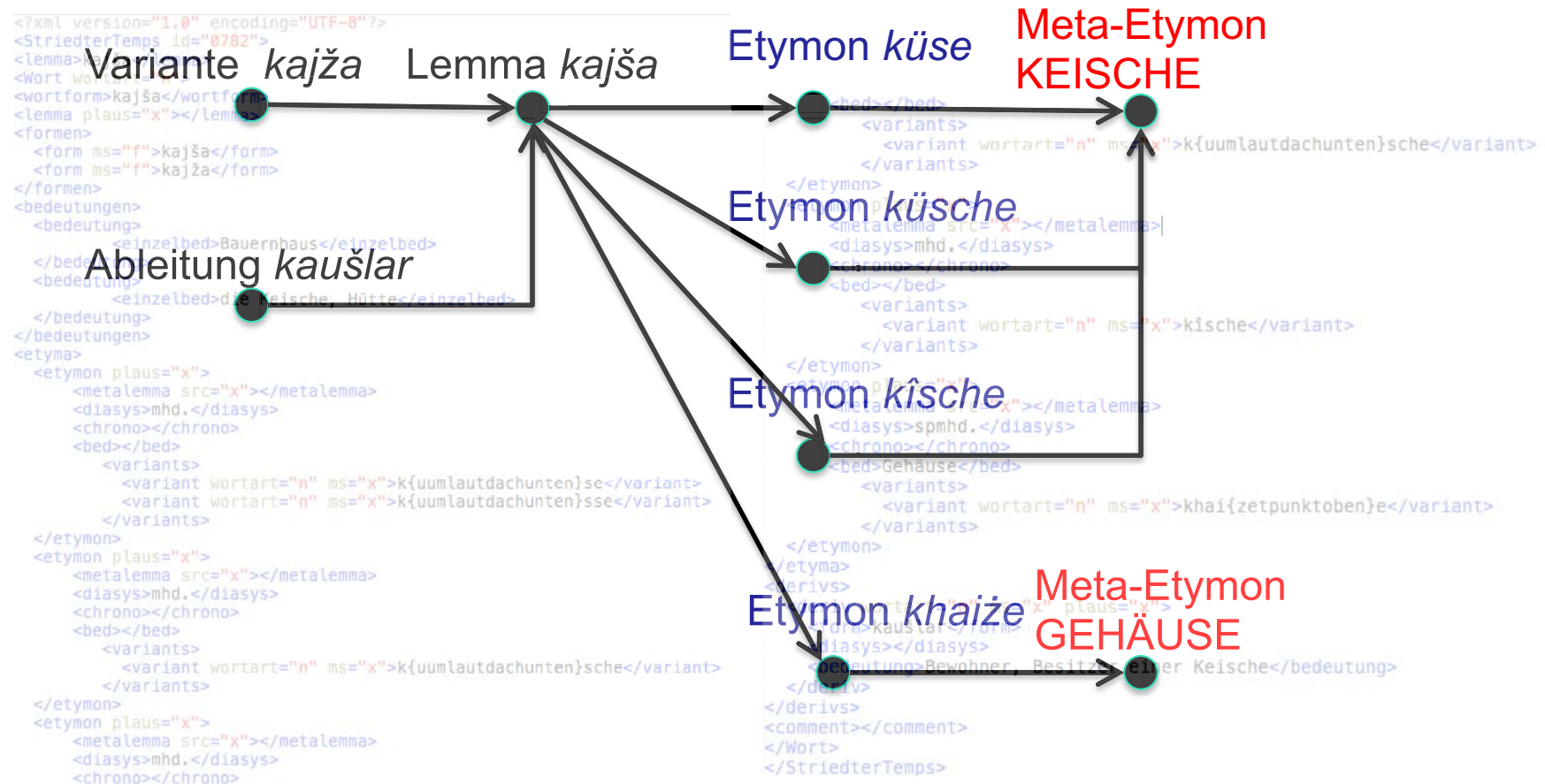
<?xml version="1.0" encoding="UTF-8"?>
<StriedterTemps id="0782">
  <lemma>kajša</lemma>
  <Wort wortart="n">
    <wortform>kajša</wortform>
    <lemma plaus="x"></lemma>
    <formen>
      <form ms="f">kajša</form>
      <form ms="f">kajža</form>
    </formen>
    <bedeutungen>
      <bedeutung>
        <einzelbed>Bauernhaus</einzelbed>
      </bedeutung>
      <bedeutung>
        <einzelbed>die Keische, Hütte</einzelbed>
      </bedeutung>
    </bedeutungen>
    <etyma>
      <etymon plaus="x">
        <metalemma src="x"></metalemma>
        <diasys>mhd.</diasys>
        <chrono></chrono>
        <bed></bed>
        <variants>
          <variant wortart="n" ms="x">k{uumlautdachunten}se</variant>
          <variant wortart="n" ms="x">k{uumlautdachunten}sse</variant>
        </variants>
      </etymon>
      <etymon plaus="x">
        <metalemma src="x"></metalemma>
        <diasys>mhd.</diasys>
        <chrono></chrono>
        <bed></bed>
        <variants>
          <variant wortart="n" ms="x">k{uumlautdachunten}sche</variant>
        </variants>
      </etymon>
      <etymon plaus="x">
        <metalemma src="x"></metalemma>
        <diasys>mhd.</diasys>
        <chrono></chrono>
      </etymon>
    </etyma>
  </Wort>
</StriedterTemps>
  
```

```

    <bed></bed>
      <variants>
        <variant wortart="n" ms="x">k{uumlautdachunten}sche</variant>
      </variants>
    </etymon>
    <etymon plaus="x">
      <metalemma src="x"></metalemma>
      <diasys>mhd.</diasys>
      <chrono></chrono>
      <bed></bed>
      <variants>
        <variant wortart="n" ms="x">kische</variant>
      </variants>
    </etymon>
    <etymon plaus="x">
      <metalemma src="x"></metalemma>
      <diasys>spmh.</diasys>
      <chrono></chrono>
      <bed>Gehäuse</bed>
      <variants>
        <variant wortart="n" ms="x">khai{zetpunktoben}e</variant>
      </variants>
    </etymon>
  </etyma>
  <derivs>
    <deriv wortart="n" ms="x" plaus="x">
      <form>kaušlar</form>
      <diasys></diasys>
      <bedeutung>Bewohner, Besitzer einer Keische</bedeutung>
    </deriv>
  </derivs>
  <comment></comment>
</Wort>
</StriedterTemps>
  
```

automatisierte Überführung der Bögen in XML-Dateien

## Teilweise manuelle Retrodigitalisierung 3



automatisierte, 'Übersetzung' lexikografischer Information aus den XML-Dokumenten in Fragmente des gerichteten azyklischen Graphs

---

# Fahrplan

- Grundsätzliche Informationen zum IDS-Projekt „Lehnwortportal“
  - technische Umsetzung des Internetportals
  - Aufbereitung eines älteren Print-Lehnwörterbuchs mit diskursiver Mikrostruktur
  - lexikografische Probleme und Fazit
-

## Lexikografische Probleme und Fazit 1

- Identifikation von Angabetypen schwierig
- grammatische Angaben fehlen häufig
- direkte Entlehnung oder innerslovenische Ableitung?
- unnötige Ambiguitäten; Interpretations- und Recherchebedarf
- irreführende Erläuterungen
- Konsultation des lautgeschichtlichen Teils erforderlich
- Entlehnungszeitraum unklar
- meist direkte Übernahme von Angaben, Transkriptionen, orthografischen Gepflogenheiten aus den verschiedenen Quellen

## Lexikografische Probleme und Fazit 2

- typische Probleme bei Striedter-Temps:
  - Entscheidungsproblem: ist *coprati* direkt **entlehnt** aus dt. *zaubern* oder **inerslovenische Ableitung** zum Lehnwort *coper* aus dt. *Zauber*?
  - Entscheidungsproblem: **Variante** oder **inerslovenische Ableitung**?  
*kragen* »Kragen, Hals«, kä.slov. (Gutsm. 162); sonst slov.: *krógæn* dass. (Graf., Razpr. I, 379); daneben begegnen noch: *krágelj* und *krôgljič* (mit slov. Suffix)
- **Mehrfachentlehnungen**: Zuordnung von slovenischen Varianten zu passenden Etymaformen aus verschiedenen Sprachstufen geht auf den Erfassungsbögen verloren
- durch die **Bilddigitalisierung** geht auf der Ebene der Einzelartikel für den menschlichen Leser keine Information verloren
- für **Suchfunktionen** im Portal ist die manuelle Extraktion von Information (Wortformen mit Bedeutungs- und grammatischen Angaben und ihre Relationen zueinander) unumgänglich
  - dies bedingt eine problematische **Komplexitätsreduktion** und nötigt häufig zu arbiträren lexikografischen Entscheidungen

# Zusätzliche Folien

## Repräsentation der Portalentitäten und ihrer Relationen

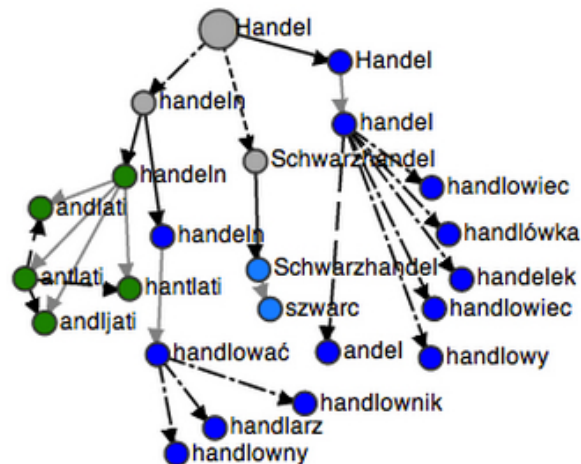
...	...	...	...	...	...	...	...	...	...	...
367	cecha	WDLP	123	var_2	„Gilde“	nom.f.	1472	nein	Lw.	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
159	367	899	Derivat	WDLP	–	w.				
...	...	...	...	...	...	...	...	...	...	...

*idealisierter Beispiel-Datensätze aus zwei relationalen Datenbanktabellen; oben: eine Portalinstanz (Wortform) mit ihren Eigenschaften; unten: Relation „ist Derivat zu“ zwischen zwei Portalinstanzen*

- zunächst **automatisierte Gewinnung** der Portalinstanzen und Vernetzungen aus den XML-Dokumenten; manuelle lexikographische Nacharbeit
- Möglichkeiten der effektiven Speicherung des gerichteten azyklischen Graphen:
  - transitiven Abschluss der Relationen in relationalen DB-Tabellen verwalten (derzeitige Lösung, siehe Abbildungen)
  - NoSQL-Graphendatenbank (z.B. neo4j)
  - Topic Map (z.B. Ontopia)

Suche

## Das Lehnwortportal Deutsch des IDS



ist ein Portal für Wörterbücher zu Entlehnungen aus dem Deutschen. Die Wörterbücher können als Einzelwerke konsultiert werden; im Portal sind sämtliche Informationen jedoch zusätzlich als komplexes, wörterbuchübergreifendes Netzwerk von Lehn- und Herkunftswörtern repräsentiert. So ist das Portal als **umgekehrtes Lehnwörterbuch** nutzbar: Im Wörterbuch der Herkunftswörter kann ausgehend von deutschen Lexemen in allen Wörterbüchern nach zugehörigen Lehnwörtern in anderen Sprachen gesucht werden. Für Experten bietet das Portal komplexe Suchfunktionen an, um Lehn- oder Herkunftswörter beispielsweise auch anhand von grammatischen Merkmalen oder zeitlich-dialektaler Zuordnung zu finden.

Herkunftswörterbuch

Herkunftswörter

Lehnwörterbücher

Polnisch

Teschener Polnisch

Slovenisch

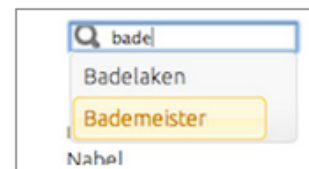
### Einzelwörterbücher

Verwenden Sie das **Wörterbuchmenü** auf der rechten Seite, um eines der Lehnwörterbücher des Portals aufzuschlagen. Das System **merkt sich**, an welcher Stelle Sie ein Wörterbuch zuletzt aufgeschlagen hatten, so dass Sie bequem zwischen verschiedenen Wörterbüchern hin- und herspringen können.

Polnisch  
Teschener Polnisch  
Slovenisch

### Stichwortsuche

Nutzen Sie in allen Wörterbüchern die **Navigationshilfen**: Auf der linken Seite steht Ihnen neben einem scrollbaren Auszug aus der Stichwortliste des gewählten Wörterbuchs auch eine schnelle **"Autocomplete"**-Suche nach dem Wortanfang des



Startseite  
des Portals  
(Stand  
September  
2012)



---

# Probleme für eine Abbildung von Wortformrelationen auf einen gerichteten azyklischen Graphen 1:

## Kontamination

slovenisch *zemnahti* ‚Mariae Verkündigung‘:

- *ebennaht* stf. ‚Tag- u. Nachtgleiche‘ (im gleichen Monat);
  - erinnert an Feiertag *ëbenwîhe* stf. ‚festum circumcisionis‘
  - letzteres ist „*sieben naht*“ von Weihnachten entfernt
-

# Probleme für eine Abbildung von Wortformrelationen auf einen gerichteten azyklischen Graphen 2:

## unklare Zuordnung von slovenischen Varianten zu Etyma bei Mehrfachentlehnung

*gūmpež* 1. »Lümmel, Tölpel«; 2. »Gimpel, Blutfink«; *gūmpec*, *-pca* 1.»Lümmel, Tölpel«; 2. »Schlag«; die Bdtg. 2. nach *gūmpati* »mit geballter Faust schlagen« (Plet. I, 260); *gumpati* im k.ä. Slovenisch »puffen« (Gutsm. 222); die angeführten Wörter lassen sich alle auf mhd. *gumpen* »springen, hüpfen« zurückführen; in Kärnten *gūmpe*, *gūmp.l* »nährischer Mensch, Grobian«, *gump-n* »tolle Sprünge machen, stoßen, sich wie ein Narr gebärden«; Bedeutungserweiterung von »springen« zu »nährisch umherspringen, nährisch, dumm sein«; auch die Bdtg. »mit geballter Faust schlagen, puffen« findet im kärntnerischen »stoßen« ihre Entsprechung. Ebenfalls zu *gumpen* gehört dt. *Gimpel*, spmhd. *gūmpel*; vgl. slov. *gūmpel* (s. dort). Slov. *gūmpež* und *gūmpec* sind mit slov. Suffixen weitergebildet. Vgl. auch Štrek., Arch. 14, 525.

*fruštuk* »Frühstück« (Meg.); im Görz. Mittelkarstdialekt: *frjštik* dass. (Štrek. Görz. 22); im k.ä. Slovenischen *frštek* neben *brštek* (Kranzmayer 12); um die Sprachinsel Zarz: *brštik* dass.; aus bair.-öster. *Früstück* »Frühstück«, mdal. *fruštuck*, auch *fruštick*, *vršštik* (Zarz) (s. Hauptkatalog), nhd. *Frühstück*, erst spmhd. belegt als *vrustücke*, *vrüestüc*. Die Formen mit *b-* sind älter, hörsprachliche Entlehnungen des Hochmittelalters (s. Kranzmayer a. a. O.). Vgl. skr. *frūštuk*, *froštuk*, *froštukalj*, *frīšteč*.