

Sinasi Derebey

Ein digitales Wörterbuch als Hilfsmittel für die deutsch-mazedonische Kommunikation

1. Vorgeschichte

Die genaue Anzahl der Mazedonisch sprechenden Menschen in Deutschland ist schwer zu ermitteln. Sie wird auf ca. 75.000-85.000 geschätzt. Die Schätzung für die Gesamtzahl der Bürger aus Mazedonien in allen deutschsprachigen Ländern beläuft sich auf ca. 200.000¹.

Sie umfasst die erste, zweite und dritte Generation von Migranten, die seit Ende der fünfziger, Anfang sechziger Jahren des vorigen Jahrhunderts nach Deutschland kamen. Die meisten von ihnen sind in der deutschen Gesellschaft wohl integriert, mit starken Bindungen zu ihrem Heimatland. In größeren Städten, unterstützt durch die Lokalverwaltungen, organisieren sie den Zusatzunterricht für ihre Kinder in Mazedonisch. Geographisch sind sie hauptsächlich um die deutschen Großstädte wie München, Berlin, Hamburg, Stuttgart und im Ruhrgebiet verteilt.

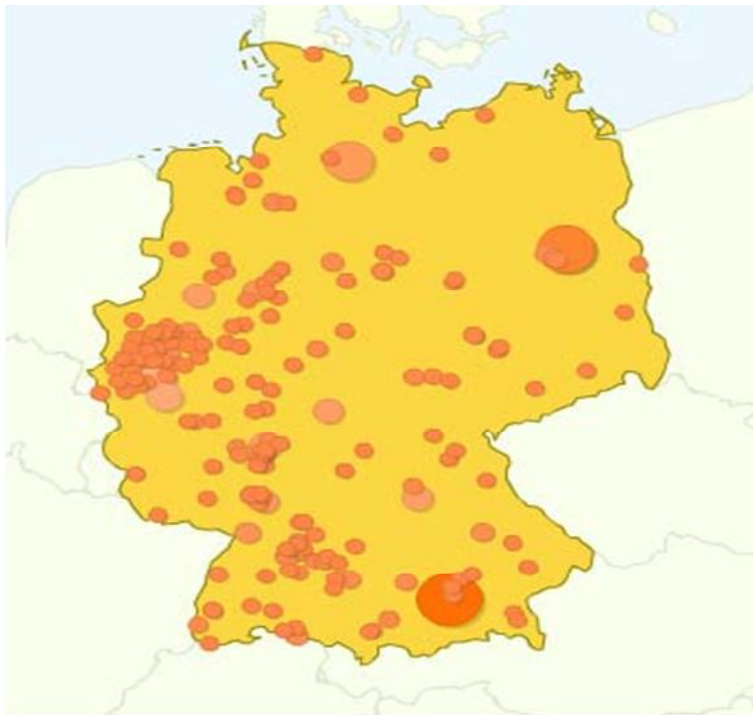


Abbildung 1: Geographische Verteilung von Mazedonisch Sprechenden in Deutschland

¹<http://de.statista.com/statistik/daten/studie/1221/umfrage/anzahl-der-auslaender-in-deutschland-nach-herkunftsland>

<http://www.utrinski.com.mk/?ItemID=2020C74B044D39409EC970182A275B19>

Die Abbildung 1 wurde mit Hilfe von Google Analytics erstellt; dazu diente die Analyse der Zugriffe auf das weiter unten beschriebene Wörterbuch

In Mazedonien lernen laut Angaben des Goethe-Instituts in Skopje ca. 50.000-60.000 Schüler Deutsch. Hauptzentren sind dabei die Städte Skopje, Bitola und Štip. Das Interesse an der deutschen Sprache ist traditionell groß, und sie wird sowohl an den Grund- als auch Mittelschulen als erste oder zweite Fremdsprache gelehrt.

In Berlin ist seit dem Jahr 2000 die Deutsch-Makedonische Gesellschaft tätig. Durch die Unterstützung der Mitglieder dieser Gesellschaft wurde das bi-direktionale deutsch-mazedonische Wörterbuch verbreitet, das im Internet unter der Webadresse www.makedonisch.info erreichbar ist. Den Link zum Wörterbuch kann man auf den Internet-Seiten der Deutschen Welle, des Goethe-Instituts und der Deutschen Botschaft in Skopje finden.

2. Zweck

Mit dem Gedanken, zunächst ein Hilfsmittel für die mazedonische Sprachgemeinschaft in Deutschland zu schaffen, wurde im Jahr 2003 begonnen, das digitale deutsch-mazedonisch-deutsche Wörterbuch aufzubauen. Die Hoffnung dabei war, die bi-lingualen Fähigkeiten der Kinder der Migranten, welche in Deutschland aufgewachsen und geschult wurden, zu mobilisieren und aufbauend auf deren Wissen ein praktisches Werkzeug zu schaffen. Das bedeutet: Ein aktuelles Wörterbuch hoher Qualität mit Erweiterbarkeit in beide Sprachrichtungen.

Ein weiteres Ziel war, die computergestützte Übersetzung zwischen den beiden Sprachen voranzutreiben. Angesichts der europäischen Perspektiven Mazedoniens ist ein solches Werkzeug zur Beschleunigung der Kommunikation mit den deutschsprachigen Ländern eminent wichtig.

Als Nebeneffekt hat das Wörterbuch im Nachhinein auch Möglichkeiten zur Sprachforschung und quantitativen Analyse der mazedonischen Sprache eröffnet.

Am 16. Mai 2007 hatte die Generalversammlung der Vereinten Nationen mit der Resolution GA/10592 das Jahr 2008 zum Internationalen Jahr der Sprachen erklärt. Das war eine Aufforderung, die kleinen Sprachen zu schützen und sie vom Aussterben zu bewahren. Das Internet ist einerseits ein nützliches Mittel zur Verbindung von kleinen Gemeinschaften, die hauptsächlich Fokussierung auf die englische Sprache macht diese jedoch zu einer neuen „Lingua Franca“, sehr zum Nachteil von anderen Sprachen.

Das deutsch-mazedonisch-deutsche Wörterbuch ist auch ein Beitrag zum Schutz und zur Förderung der mazedonischen Sprache.

3. Konzept

Internet-Wörterbücher als relativ neue Errungenschaft bieten entscheidende Vorteile gegenüber den klassischen, gedruckten Wörterbüchern: Nicht nur wegen der höheren Geschwindigkeit der Veröffentlichung und der Möglichkeit, vorhandene Ressourcen im Internet einzubinden, sondern auch wegen der Möglichkeit der Interaktion mit dem Nutzer.

Das Publizieren eines gedruckten Wörterbuches ist eine mühevollere Angelegenheit, und es nimmt viel Zeit in Anspruch. Die erste Ausgabe enthält ca. 4-8% Fehler und es dauert mindestens ein Jahr, um das Wörterbuch zu veröffentlichen. Damit eine annehmbare Qualität erreicht wird (Fehlerquote unter 1%), benötigt man mindestens drei oder vier Auflagen. Bestenfalls benötigt man für ein gutes Wörterbuch zehn Jahre mühevollere Arbeit. Das macht es sehr teuer, besonders wenn die Anzahl der Verkaufsexemplare klein ist.

Im Vergleich dazu wird das deutsch-mazedonisch-deutsche Internet-Wörterbuch einmal monatlich aktualisiert. Die Korrektur und Eingabe neuer Wörter dauert nur Sekunden.

Eine weitere Funktionalität des Internet-Wörterbuches ist die Möglichkeit der schnellen Rückkopplung mit den Nutzern. Sie berichten sofort über fehlende Einträge, fehlerhafte Wörter, falsche oder bessere Übersetzungen oder über Probleme bei der Nutzung. Ähnliches ist beim gedruckten Wörterbuch fast unmöglich.

Dieses Internet-Wörterbuch verbindet vorhandene Ressourcen im Internet und gibt dem Nutzer die Möglichkeit, sein Wissen zu vertiefen. Eine solche Verbindung auf der deutschen Seite besteht zum „Digitalen Wörterbuch der Deutschen Sprache“ (www.dwds.de) der Berlin-Brandenburgischen Akademie der Wissenschaften. Auf der Gegenseite ist das Digitale Wörterbuch der Mazedonischen Sprache (www.makedonski.info), das sich auf das Wörterbuch der mazedonischen Sprache von Blaže Koneski stützt.

Ein gedrucktes Wörterbuch ist eine sortierte Liste von Wörtern (Lexemen), in der jedem Wort der „Quellsprache“ ein oder mehrere Wörter der „Zielsprache“ zugeordnet sind.

Der Suchalgorithmus ist denkbar einfach: Dadurch dass die Einträge in alphabetischer Reihenfolge sortiert sind (im Deutschen A-Z), wird ein deutsches Wort durch Blättern gefunden und dann der mazedonische Eintrag loziert.

Die Suche in die andere Richtung ist nicht möglich. Dazu wird ein anderes, „umgekehrtes“ Wörterbuch benötigt. Um ein mazedonisches Wort zu finden, benötigt man jetzt ein Wörterbuch, welches nach der Reihenfolge des mazedonischen Alphabets sortiert ist (A-III). Da die mazedonische Sprache das kyrillische Alphabet benutzt, ist die Sortierfolge in den beiden Wörterbüchern unterschiedlich.

*Das bedeutet: Das **Deutsch-Mazedonische** Wörterbuch ist sortiert in der Reihenfolge des deutschen Alphabets (lateinisch) und das **Mazedonisch-Deutsche** Wörterbuch in der Reihenfolge des mazedonischen Alphabets (kyrillisch).*

Ein Computer benötigt keine Sortierfolge zur Suche. Im Gegensatz zum menschlichen Leser, kann der Computer durch seine hohe Geschwindigkeit jedes Wort durch eine Komplettsuche des Wörterbuches finden.

DG Aal <i>m</i> BIO (geräucherter Aal)	јагула (чадена јагула)
DG aaen, sich <i>v</i>	(се излежува испружен)
DG Aalfang <i>m</i>	лов на јагули
DG aalglatt <i>adj</i>	(лизгав како јагула)
DG Aar <i>m</i> BIO] [ALT	ореп
DG Adler <i>m</i> BIO	ореп
DG Aas <i>n</i>	мрша; труп
DG aasen <i>v</i>	се расфрпа; (живее расипнички)
DG Aasgeier <i>m</i> BIO	мршојад; мршојадец
D ab <i>adv</i> (ab morgen wird gefastet)	од (од утре се пости)
D ab <i>prep</i> (auf und ab)	од (горе – долу)
D ab <i>prep</i> (ab und zu)	од (одвреме – навреме)
D ab <i>adv</i> (ab nach Hause!)	право (право дома)
D ab <i>adv</i> MIL (Gewehr ab!)	долу (пушката долу)
DG Aba <i>f</i>	(арапски долг капут)
DG abändern <i>v</i>	измени; изменува
DG abänderlich <i>adj</i>	променлив
DG Abänderung <i>f</i>	промена; измена
DG Abänderung <i>f</i>	преиначување
DG Abänderungsvorschlag <i>f</i>	(предлог за промена); предлог-амандман

Abbildung 2: Deutsch-Mazedonisches Wörterbuch, Einträge sortiert nach dem deutschen Alphabet

Jedem deutschen Wort können ein oder mehrere mazedonische Wörter zugeordnet sein (mehrere Bedeutungen, Synonyme, usw.) und umgekehrt.

Es gibt auch Fälle, in denen kein adäquates Wort als Übersetzung zur Verfügung steht. Dann besteht die Übersetzung aus einer Umschreibung in Klammern.

Das erklärt, warum ein digitales Wörterbuch, obwohl es keiner Sortierfolge bedarf, eine „Sprachrichtung“ benötigt. Das ist dann besonders notwendig, wenn die Unterschiede im Wortschatz beider Sprachen groß sind, wie z.B. im Falle der Kombination Deutsch / Mazedonisch.

(Beispiel: Das Wort „Ossi“ lässt sich nicht wortwörtlich übersetzen, deshalb wird dieses Wort in einem Klammerausdruck umschrieben. Das gilt umgekehrt auch für das mazedonische Wort „домазет“ (jemand der einheiratet). Das bedeutet, dass das Wort „домазет“ nur im **mazedonisch**-deutschen und das Wort „Ossi“ nur im **deutsch**-mazedonischen Wörterbuch vorhanden ist.)

Was ist neu an diesem Wörterbuch?

Beide Wörterbücher sind in einem Wörterbuch integriert. Bei der Suche werden beide Sprachrichtungen durchsucht und die Treffer angezeigt.

Jedes Wort auf der deutschen Seite ist mit dem „Digitalen Wörterbuch der Deutschen Sprache“ (DWDS) verlinkt und der Nutzer kann durch einen Klick dort weitere Informationen abrufen. Ebenso besteht die Möglichkeit, sich auf der mazedonischen Seite mit dem Digitalen Wörterbuch der mazedonischen Sprache („Tolkovnik“) zu verbinden.

Diese Verbindungen werden durch kleine farbige Symbole angezeigt. Ein rotes „**T**“ zeigt den Link zum „Tolkovnik“ und ein „**D**“ zum DWDS.

Makedonisch-Deutsches Wörterbuch

T шеќер <i>m</i> ESSG	Zucker
T шеќерка <i>f</i>	Apfelsorte
T шеќерлија <i>adj</i>	kandiert; zuckersüß
T шеќерлиса, се <i>v</i>	verzuckern
T шеќерче <i>n</i>	Bonbon; Zuckerchen

Abbildung 3: Symbole mit Verlinkung

Durch den Klick auf das rote „**T**“ im ersten Wort („шеќер“) wird der nachfolgende Eintrag aus dem „Tolkovnik“ angezeigt:

шеќер
Необработен текст
шеќер <i>m.</i> сахароза, засладувач во вид на ситни кристали; чај-шеќер (разг.) шеќер во коцки; ситен шеќер; шеќер во прав; жолт шеќер; фабрика за шеќер; сладок како шеќер.

Abbildung 4: Eintrag aus dem „Tolkovnik“ nach dem Klick auf „шеќер“

Neue Wörter werden durch die Rückkopplung von den Benutzern oder durch periodischen automatischen Abgleich mit Wortbeständen aus dem Internet in das Wörterbuch aufgenommen.

4. Herausforderungen

Das vollständige mazedonische Wörterbuch von Koneski („Правописен речник на македонскиот јазик“) zählt ca. 42.000 Einträge. Das deutsche Universalwörterbuch von Duden² enthält ca. 250.000 Wörter. Das bedeutet, dass der Unterschied im Wortbestand zwischen den beiden Sprachen groß ist. Dieses führt auch zu einer relativ hohen Anzahl der Wörter, die umschrieben werden.

Die kyrillische Schrift mit den mazedonischen Sonderzeichen führt manchmal zu Schwierigkeiten, da einige syntaktische Zeichen (z.B. Apostroph in „’рж“) auch als Inhaltsträger gelten.

Das Fehlen eines repräsentativen Textkorpus der mazedonischen Sprache bereitet weitere Schwierigkeiten. Das wirkt sich auf die automatische Suche und den Abgleich des mazedonischen Wortbestandes aus. Deshalb werden gegenwärtig die Identifikation und Erkennung von fehlenden Einträgen manuell durchgeführt. Das ist, wie man sich vorstellen kann, aufwendig und teuer. In diesem Zusammenhang muss man auch auf die komplizierte und unklare Rechtesituation bezüglich der Copyrights bei der Nutzung von Texten für die Textkorpora in Mazedonien besonders hinweisen.

Sowohl die deutschen als auch die mazedonischen Wörter werden intern nach verschiedenen Merkmalen klassifiziert. Diese Merkmale erlauben eine spätere Auswertung nach der grammatikalischen Kategorie (Nomen, Verb, Adjektiv, Adverb usw.), Etymologie (Herkunft und Verbindung zu Wörtern aus anderen Sprachen), semantischem Kontext (Medizin, Justiz, usw.) und dem Modus (veraltete Begriffe, Dialekt, Pejorativ usw.). Dieser Prozess ist noch nicht abgeschlossen.

5. Status

Das digitale deutsch-mazedonische Wörterbuch beinhaltet gegenwärtig ca. 120.000 Lexeme (ca. 80.000 im deutschen und ca. 40.000 im mazedonischen Bereich). Es wird monatlich ca. 100.000-mal von etwa 5.000 eindeutigen Besuchern benutzt.

Es ist unter den Adressen www.makedonisch.info und www.germanski.com.mk ständig im Internet frei verfügbar. Es wird auf einer Infrastruktur von 3 Servern unterbrechungsfrei, 24 Stunden täglich, betrieben.

Die Rückkopplung ist durchaus positiv, besonders von den Besuchern aus dem universitären Bereich.

² http://www.duden-suche.de/werke/werk_info.php?shortname=fx

1.	ruprecht-karls-universitaet heidelberg
2.	technische universitaet dortmund
3.	universitaet muenster
4.	heinrich-heine-universitaet duesseldorf
5.	martin-luther-universitaet halle-wittenberg
6.	bauhaus-universitaet weimar scc
7.	freie universitaet berlin
8.	ludwig-maximilians-universitaet muenchen
9.	ruhr-universitaet bochum
10.	universitaet bonn

Abbildung 5: Nutzer aus dem Uni-Bereich

6. Ausblick

Ein digitales Wörterbuch ist der erste Schritt in Richtung „computergestützte Übersetzung“. Das bisher von Google benutzte Verfahren der statistischen Übersetzung wird im Falle von Deutsch / Mazedonisch nicht ganz funktionieren, weil der Unterschied im Wortschatz sehr groß ist und der vorhandene Bestand an zuverlässigen Übersetzungen zwischen Deutsch und Mazedonisch klein ist. Deshalb erscheint die Wörterbuch-gestützte Übersetzung als der einzig vernünftige Weg.

Eine weitere Aufgabe liegt in der Implementierung der Lemmatisierungsregeln für die mazedonische Sprache. Die Rückführung eines Wortes aus einer beliebigen morphologischen Form in die Nominalform ist die Voraussetzung für den Zugriff auf das Wörterbuch. Diese Aufgabe ist in Planung und soll demnächst implementiert werden.

Weiterhin soll der Bestand des „Tolkovnik“ komplettiert und die Qualität der Einträge verbessert werden. Dazu steht auch die Software des DWDS zur Verfügung, um ähnliche Funktionalitäten für Mazedonisch bereit zu stellen.

Auf der Wunschliste steht auch die Schaffung eines mazedonischen Textkorpus.

Der Übersetzungsprozess kann schrittweise implementiert werden: Zunächst in Form von Rohübersetzungen mit „Suchen und Ersetzen“ von Wörtern und später durch den Einsatz von grammatikalischen Regeln. Es ist auch eine Mischung aus statistischen Übersetzungsverfahren mit Wörterbuch-gestützten Verfahren denkbar.

Es sollen weitere Internet-Ressourcen eingebunden werden wie z. B. das IATE-Terminologie-[Wörterbuch der Europäischen Union \(http://iate.europa.eu\)](http://iate.europa.eu). Eine Genehmigung der zuständigen Fachgruppe der EU zur Verlinkung von IATE mit „makedonisch.info“ liegt bereits vor.

Durch Analyse etymologischer Merkmale von mazedonischen Wörtern in diesem Wörterbuch kann man feststellen, wie viele Bindungen zu den anderen Sprachen bestehen. Beispielsweise haben viele deutsche Wörter Eingang in die mazedonische Sprache gefunden, teilweise unter Modifikation der ursprünglichen Bedeutung. Im Mazedonischen hat das Wort „гелендер“ (Geländer) die gleiche Bedeutung, ebenso wie das Wort „ауспух“ (Auspuff) zum Unterschied von „трапав“ (Trauerflor).

Diese Gemeinsamkeiten könnten bei der Ausgestaltung eines Curriculums für die mazedonisch/deutsch Lernenden nützlich werden.

Zusammenfassung

Das digitale deutsch-mazedonisch-deutsche Wörterbuch ist zu einem wichtigen Werkzeug gewachsen und trägt somit der Bedeutung der deutschen Sprache in Mazedonien Rechnung. Es dient auch der mazedonischen Sprachgemeinschaft in den deutschsprachigen Ländern, um die Sprache ihres Heimatlandes zu pflegen.

Literatur:

Doneva, Ana (2007): Praktische Benutzung digitaler Wörterbücher, Sprachenpaar Makedonisch-Deutsch, Justus-Liebig-Universität Gießen, Masterstudiengang Sprachtechnologie, Betreuer: Prof. Dr. Dietmar Rösler