

AUF DEM WEG ZU EINER KARTOGRAPHIE: AUTOMATISCHE UND MANUELLE ANALYSEN AM BEISPIEL DES KORPUS ISW

Die Autorin ist Juniorprofessorin für Germanistische Linguistik an der Universität Mailand (Lingue e Letterature Straniere).

Der Autor ist wissenschaftlicher Mitarbeiter im Programmbereich „Mündliche Korpora“ in der Abteilung Pragmatik am Leibniz-Institut für Deutsche Sprache, Mannheim.

Einleitung

Orte – sowohl im Sinne von geographischen Bestimmungen als auch von ‚sozialen Räumen‘ und ‚Plätzen‘ einer Stadt – haben eine wichtige Bedeutung im sogenannten ‚Israelkorpus‘, das in Zusammenarbeit mit der Korpus-Urheberin Anne Betten von Anfang an am IDS archiviert, digitalisiert und für die Internetpräsentation bearbeitet wurde; es besteht derzeit aus 283 „Ereignissen“ (mit wenigen Ausnahmen spontanen narrativen autobiographischen Interviews). Die Interviews wurden von Anne Betten und verschiedenen temporären Mitarbeiter/innen seit 1989 in Israel mit Emigranten und Emigrantinnen aus deutschsprachigen Regionen Mitteleuropas („Jeckes“) und deren Kindern aufgenommen (vgl. die Korpora IS, ISW, ISZ¹ in der Datenbank für Gesprochenes Deutsch, DGD, Abb. 1).

ORTE HABEN IN DEN BIOGRAPHISCHEN INTERVIEWS DES ‚ISRAELKORPUS‘ EINE WICHTIGE BEDEUTUNG

Das Ziel des ursprünglichen DFG-Projekts war es zunächst, die Sprachsituation und die Gründe für die Bewahrung des Deutschen sowie die kulturelle Identität der Emigrant/innen 50 bis 60 Jahre nach ihrer Immigration in Palästina/Israel (Korpus IS, ISW) zu dokumentieren und zu erforschen; seit 1999 kamen Aufnahmen mit der 2. Generation hinzu (Korpus ISZ). Bevorzugte Themen in den Korpora IS und ISW sind u. a. Kindheit und Jugend in Deutschland/Österreich, Erfahrungen

von Antisemitismus, Flucht / Emigration, Neuanfang und kulturelle Umorientierung, während das Korpus ISZ von den Fragen ausgeht, wie sich die Interviewten als Kinder deutschsprachiger Juden gefühlt haben und wie sich das Heranwachsen in zwei Kulturen auf ihre Identitätsbestimmung auswirkte.²

Orte sind Befestigungspunkte für Gedächtnisinhalte und haben eine wichtige Funktion innerhalb der Erinnerungsarbeit (vgl. Casey 1993, S. 171). Sie verschränken sich oft mit der emotionalen Dimension: Nicht nur alte Emotionen werden wachgerufen, sondern es entstehen auch neue (Leonardi/Thüne/Betten (Hg.) 2016, S. 2). Diese können sowohl explizit thematisiert als auch nur indirekt ausgedrückt werden. Da Orte von Erlebnissen der jeweiligen Menschen abhängig sind (Becker 2019), kann sich ihre Bedeutung im Laufe der Zeit ändern (vgl. Schwitalla i. Ersch.). Für viele Interviewte des ISW-Korpus ist beispielsweise Wien mit ambivalenten Gefühlen assoziiert. Positive Emotionen, wie Liebe und Freude, werden nicht nur benannt, sondern auch beschrieben, wenn über die Kindheit gesprochen wird. Nach den ersten antisemitischen Erfahrungen oder spätestens nach dem „Anschluss“ und nach der Emigration wird jedoch die Stadt für viele Interviewte eine Stadt der Angst oder des Hasses (vgl. Brambilla/Flinz 2019).

ORTE VERSCHRÄNKEN SICH MIT DER EMOTIONALEN DIMENSION UND DER ERINNERUNGSARBEIT

Im Folgenden möchten wir unsere Pilotstudie zur Auffindung von Ortsnennungen im Korpus ISW vorstellen, die als Vorarbeit für die Erstellung einer Kartographie (vgl. § 2) dienen soll, die zur Rekonstruktion der ‚Lebenswelten‘ der zur Auswanderung gezwungenen deutschsprachigen Juden beitragen soll (vgl. <https://kartografiedesisraelkorpus.wordpress.com/projekt/>). Ziel der anvisierten Kartographie ist, die relevantesten Orte des Korpus graphisch darzustellen. Diese Visualisierung setzt aber eine Datenerfassung, Bearbeitung und Analyse voraus. Die verschiedenen Arbeitsschritte umfassen die jeweilige genaue Verortung der Erzählung in Raum und Zeit, Hinweise zu Ereignissen und zu den Sprechern im Rahmen des Korpus und zu deren in den Interviews erwähnten Familienangehörigen und Bekannten sowie Verlinkungen zur DGD. In dieser ersten Projektphase, die wir mit diesem Beitrag vorstellen möchten, geht es um die Auffindung der Orte, die vor, während und nach der Emigration im Leben der In-

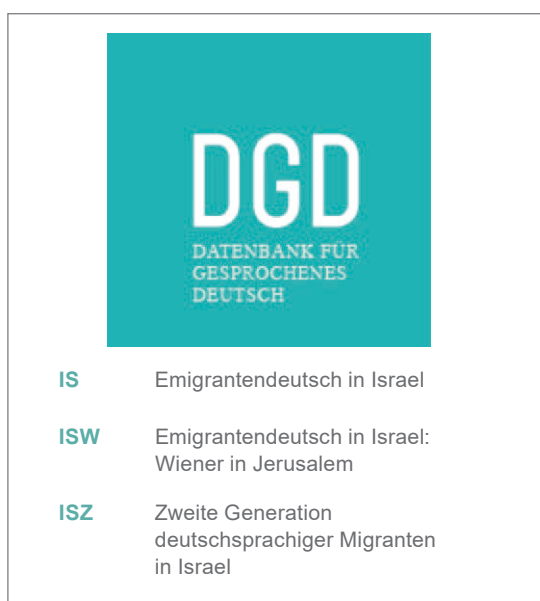


Abb. 1: Die Israelkorpora in der DGD



Abb. 2: Cover von *Wir sind die letzten. Fragt uns aus*

terviewten eine Rolle spielen. Diese sollen im Korpus herausgefiltert werden; dabei haben wir sowohl automatische als auch manuelle Ansätze benutzt, deren Vorteile und Probleme im Folgenden diskutiert werden sollen.

Orte, Erinnerungen und Emotionen

Das ‚Israelkorpus‘ wurde bereits unter den verschiedensten wissenschaftlichen Fragestellungen erforscht, die einzelne Interviews oder kleine Gruppen von Interviews fokussiert haben: Grammatische (syntaktisch-stilistische) Analysen und soziolinguistische Erklärungsansätze bestimmten die Publikationen im Zeitraum 1995-2005 (für Transkripte bzw. Auszüge aus den Interviews siehe v. a. Betten (Hg.) 1995; Betten/Du-nour (Hg.) 2000, 2004, siehe Abb. 2).

EINE PILOTSTUDIE ZUR AUFFINDUNG VON ORTSNENNUNGEN ALS VORARBEIT FÜR DIE ERSTELLUNG EINER KARTOGRAPHIE

In den daran anschließenden Arbeiten stand vor allem die dialoglinguistische und erzähltheoretische Perspektive im Zentrum (vgl. u. a. Betten 2007; Thüne 2009; Schwitalla 2011; Thüne/Leonardi 2011; Leonardi/Thüne/Betten (Hg.) 2016), während in den neuesten Untersuchungen Schwerpunkte hinzugekommen sind, die sowohl thematisch als auch methodologisch neue Wege gehen, denn die Daten werden als Ganzes in den Blick genommen (vgl. u. a. Flinz 2019; Brambilla/Flinz 2019; Flinz/Moroni 2020). Besonders im Fokus stehen dabei



Abb. 3: *Eine Kartographie des Israelkorpus*

Orte und ihr Wechselspiel mit Erinnerungen und Emotionen, wie in dem vom IISG (*Istituto Italiano di Studi Germanici*) geförderten Forschungsprojekt *Orte und Erinnerung: Eine Kartografie des Israelkorpus* (dessen zentrale Ergebnisse in Leonardi et al. (Hg.) (i. Ersch.) gesammelt sind). Als weitere Zielsetzung sollen Orte auch in kartographischer Darstellung aufgezeichnet werden (Abb. 3):

In einer ersten Pilotstudie haben wir die Erkennung von Ortsnennungen im ISW-Korpus untersucht.³ Dabei wurde einerseits automatisch mit Hilfe von *out of the box*-Ansätzen experimentiert, und andererseits wurde das Korpus manuell annotiert (vgl. § 4). Der Abgleich der automatischen Annotationen mit den manuellen zeigt, dass ein Zusammenspiel zwischen automatischen und manuellen Verfahren unumgänglich ist, da die automatischen Verfahren Ortsnennungen im relevanten Sinne nicht mit ausreichender Zuverlässigkeit identifizieren können und durch manuelle Kontextanalysen komplettiert werden müssen.

Das Korpus ISW

Das Korpus *Emigrantendeutsch in Israel: Wiener in Jerusalem* wurde größtenteils im Dezember 2008 bei einer Israel-Exkursion der Salzburger Germanistik zur Ergänzung des Korpus IS aufgenommen (siehe Tab. 1):

DAS ISW ALS UNTERSUCHUNGSKORPUS: AUTOMATISCHE *OUT OF THE BOX*-ANSÄTZE UND MANUELLE ANNOTATIONEN

Die im IDS archivierte Version des Korpus umfasst 28 Tonaufnahmen (51h und 09 min), 20 Transkripte (ca. 444.564 *tokens*, vgl. Abb. 4) nach überwiegend orthographisch erstellten Konventionen mit einigen zusätzlichen Notationen und Zusatzmaterialien wie Transkriptionskonventionen, Inhaltsangaben, Wortlisten, Lemmalisten, die mit den unterschiedlichen Funktionalitäten der DGD (vgl. dazu Deppermann/Schmidt 2014) recherchierbar sind.

Audio- und Video-Aufnahmen	Transkripte: korrigiert und verfügbar	Transkripte: unkorrigiert, auf Anfrage	Zeitraum	Ort
28 Audioaufnahmen	20	8	1998-2011	Jerusalem, Wien (1), Salzburg (2)

Tab. 1: Informationen zum ISW-Korpus

KORPUS	EREIGNISSE	SPRECHER	TRANSKRIPTE	AUDIO	ZUSATZMATERIALIEN
#	Ereignis-ID ▲ ▼	Sprechername ▲ ▼	Ortsname ▲ ▼	Erhebungsdatum ▲ ▼	
1	ISW_E_00001 ▶	Alexander, Mirjam u. Aaron	Jerusalem	1998	
2	ISW_E_00002 ▶	Baumann, Dr. Shaul u. Hanna	Jerusalem	1998	
3	ISW_E_00003 ▶	Beer, Paul Rudolf	Jerusalem	1998	
4	ISW_E_00004 ▶	Beer, Shoshana	Jerusalem	1998	
5	ISW_E_00005 ▶	Blitzer, Tova	Jerusalem	1998	
6	ISW_E_00006 ▶	Friedländer, Max	Jerusalem	1998	
7	ISW_E_00007 ▶	Goldschmidt, Anitta	Jerusalem	1998	
8	ISW_E_00008 ▶	Goldstein, Erich	Jerusalem	1998	
9	ISW_E_00009 ▶	Goldstein, Jeanette	Jerusalem	1998	
10	ISW_E_00010 ▶	Halevy, Margit	Jerusalem	1998	
11	ISW_E_00011 ▶	Hoffer, Gerda	Jerusalem	1998	
12	ISW_E_00012 ▶	Hoffer, Gerda	Jerusalem	1998	
13	ISW_E_00013 ▶	Hübner, Jehudith	Jerusalem	1998	
14	ISW_E_00014 ▶	Iskin, Charlotte	Jerusalem	1998	
15	ISW_E_00015 ▶	Jaul, Naftali	Jerusalem	1998	
16	ISW_E_00016 ▶	Karniel, Jeshajahu u. Chava	Jerusalem	1998	
17	ISW_E_00017 ▶	Lang, Ora	Jerusalem	1998	
18	ISW_E_00018 ▶	Patai, Alisa	Jerusalem	1998	
19	ISW_E_00019 ▶	Rath, Ari	Jerusalem	1998	
20	ISW_E_00020 ▶	Rath, Ari	Salzburg	1999	
21	ISW_E_00021 ▶	Rath, Ari	Jerusalem	1999	
22	ISW_E_00022 ▶	Rath, Ari	Salzburg	2000	
23	ISW_E_00023 ▶	Rosenkranz, Prof. Herbert	Jerusalem	1998	
24	ISW_E_00024 ▶	Shaked, Prof. Dr. Gershon	Jerusalem	1998	
25	ISW_E_00025 ▶	Shanan, Joel	Jerusalem	1998	
26	ISW_E_00026 ▶	Goldstein, Jeanette	Jerusalem	2011	
27	ISW_E_00027 ▶	Hoffer, Gerda	Jerusalem	2011	
28	ISW_E_00028 ▶	Rath, Ari	Wien	2010	

Abb. 4: Screenshot der ISW-Ereignisse in der DGD



Abb. 5: Fotocollage einiger ISW-Interviewpartner⁴

Es handelt sich um 24 jüdische Frauen und Männer im Alter von 69 bis 90 Jahren, die meist in Wien geboren und/oder aufgewachsen sind und die nach dem ‚Anschluss‘ größtenteils ohne Eltern mit der Jugendliaja Österreich verlassen haben: drei Paare (Mirjam und Aaron Alexander; Dr. Shaul Otto und Hanna Baumann; Jeshajahu und Chava Karniel), zehn Frauen und acht Männer (siehe Abb. 5).

id	token	norm	lemma	pos	GN	CUSTOM	GEO	NER	NEL
w1157	war	war	sein	VAFIN					
w1158	war	war	sein	VAFIN					
w1159	der	der	die	ART					
/									
w1160	der	der	die	ART					
w1161	Hitler	Hitler	Hitler	NE					http://de.wikipedia.org/wiki/Adolf_Hitler
w1162	ist	ist	sein	VAFIN					
w1163	doch	doch	doch	ADV					
w1164	gekommen	gekommen	kommen	VVPP					
w1165	in	in	in	APPR					
w1166	Deutschland	Deutschland	Deutschland	NE	B-ORT	B-ORT		B-LOC	http://de.wikipedia.org/wiki/Deutschland
w1167	neunzehnhun	neunzehnhun	neunzehnhun	ADJD					
(-)									
.									
w1168	Mit	Mit	mit	APPR					
w1169	dem	dem	die	ART					
w1170	Beginn	Beginn	Beginn	NN	B-ORT				
w1171	dort	dort	dort	ADV					
,									
(
...									
)									
w1172	i	i	i	ADJA					
/									
w1173	in	in	in	APPR					
w1174	Deutschland	Deutschland	Deutschland	NE	B-ORT	B-ORT		B-LOC	http://de.wikipedia.org/wiki/Deutschland
(-)									
w1175	konnte	konnte	können	VMFIN					
w1176	man	man	man	PIS					
w1177	noch	noch	noch	ADV					
w1178	leben	leben	leben	VVINFIN					
.									
w1179	Es	Es	es	PPER					
w1180	war	war	sein	VAFIN					
w1181	in	in	in	APPR					
w1182	Wirklichkeit	Wirklichkeit	Wirklichkeit	NN	B-ORT				
w1183	von	von	von	APPR					
w1184	von	von	von	APPR					
w1185	dem	dem	die	ART					
w1186	Deutschland	Deutschland	Deutschland	NE	B-ORT	B-ORT		B-LOC	http://de.wikipedia.org/wiki/Deutschland

Abb. 6: Screenshot des automatischen Annotationsansatzes im Ereignis ISW_E_00015 (T_02)

DAS KORPUS ISW: 28 TONAUFNAHMEN, 20 TRANSKRIPTE UND ZUSÄTZLICHE MATERIALIEN

Die narrativen Interviews konzentrieren sich auf die Biographie der Sprecher vor, während und nach der Emigration. Wiederkehrende Themen sind u. a. Kindheit und Jugend in Österreich, Erfahrungen von Antisemitismus, Flucht/ Emigration, Neuanfang und kulturelle Umorientierung.

Auf dem Weg einer Kartographie: automatische und manuelle Analysen

Um die Ortsnennungen im ISW-Korpus zu identifizieren, wurden, wie oben bereits ausgeführt, zwei Arten von Annotationen durchgeführt: automatische und manuelle. Zunächst wurden automatisch die Wörter, deren Lemmata in GermaNet⁵

Kategorien NER	Identifizierte Orte	Fehlerhaft identifizierte Orte
B-LOC	Österreich (19), Neunkirchen (13), Deutschland (9), Wien (8), Israel (4), Niederösterreich (4), Wiener Neustadt (3), Palästina (3), Ungarn (3), Dachau (2), die Schweiz (2), Polen (2), Schönbrunn (2), Sopron (2), Burgenland (1), Europa (1), Göringstadt (1), Neusiedlersee (1), Nordamerika (1), Ödenbrug (1), Salzburg (1), Südamerika (1)	Dollfuß (1), LACHEN (1), Göring (1), Starhemberg (1)
B-ORG	Humanistisches Gymnasium (1), Roten Falken (1), SS (1), Volksschule (1)	
B-PER	Hitler (7), Schuschnigg (5), Jabotinsky (2), Seipel (1), Starhemberg (1)	Israel (1), Jud (1), Kippa (1)

Tab. 2: Annotierte Lexeme auf der Basis des NER-Systems

(Hamp/Feldweg 1997) in mindestens einer Bedeutung mit dem Bedeutungsfeld ‚Ort‘ assoziiert sind, als Orte annotiert (Spalte GN).

Dann wurden alle Wörter automatisch in der GeoNames-Datenbank nachgeschlagen und bei vorhandenem Eintrag die assoziierten Koordinaten übernommen (Spalte GEO). Anschließend wurden mit Hilfe des Flair Named Entity Taggers (Akbik et al. 2018) Wörter und Phrasen automatisch mit den vier klassischen Kategorien von Eigennamen PER, LOC, ORG und MISC annotiert. Manuell wurden auch Vorkommen von Zeitangaben (TIME) annotiert. Abschließend wurden alle Wörter und Phrasen, deren Referenten einen Eintrag in Wikipedia besitzen (Spalte NEL), annotiert (Abb. 6).⁶

ORTSNENNUNGEN WURDEN AUTOMATISCH UND MANUELL ANNOTIERT

Parallel dazu wurde das ISW-Korpus komplett von einem Annotator A (Spalte CUSTOM) und partiell (eine Auswahl an fünf Transkripten) von einem Annotator B manuell annotiert.⁷

Um einen Einblick in die Problematik zu geben, soll im Folgenden stichprobenartig ein Abschnitt (T_02) eines Interviews besprochen werden (ISW_E_00015⁸), dessen automatische Annotation exemplarisch erläutert wird.

Die Ergebnisse (vgl. Abb. 7) zeigen, dass von 2520 im Abschnitt automatisch getaggten Wörtern 108 auf der Basis von GermaNet als Orte kategorisiert worden sind, 8 wurden als Orte geolokalisiert, 88 als Orte mit NER identifiziert, und bei 88 wurden Wikipedia-Links eingefügt.

Bei einer genaueren Betrachtung sind folgende Ungenauigkeiten oder Fehler aufgefallen:

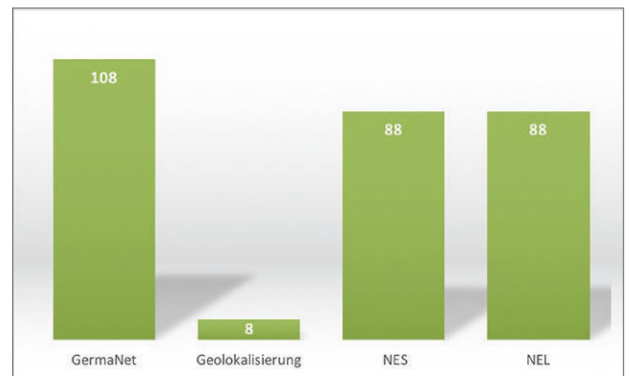


Abb. 7: Ergebnisse der automatischen Annotation

- Bei GermaNet sind 108 Wörter als Orte klassifiziert worden, darunter befinden sich Länder und Regionen (wie *Niederösterreich, Österreich, Palästina, Israel, Europa, die Schweiz, Südamerika, Nordamerika, Polen, Deutschland, Ungarn, Burgenland, Libanon*), Städte oder Stadtteile (wie *Wiener Neustadt, Wien, Dachau, Salzburg*), allgemeine Ortsnennungen (*Industriestadt, Land, Staat, Grenze, Platz, Nachbarland, Stadt, Ort, Zentrum*) und Orte innerhalb einer Stadt (*Straße, Haus, Gefängnis, Lager, Konzentrationslager, Friedhof, Grab*). Als ‚Orte‘ (B-ORT) wurden jedoch auch Lexeme markiert, die nur unter bestimmten Bedeutungsfacetten als Ort gelten können, aber nicht in unserem Kontext (wie z.B. ‚Weg‘ in *auf illegalem Weg*, ‚Grund‘ in *ohne jeden Grund*; ‚Nähe‘ in *in der Nähe* und ‚Umgebung‘ in *in einer feindlichen Umgebung*) oder Lexeme, die nur im weiteren Sinne mit Ortsangaben verbunden sind (wie *Seite, Richtung*) oder Lexeme, deren Verbindung zu Ortsangaben kaum verständlich ist (wie *Mathe, Mathematik, Niveau, Beginn, Wirklichkeit, Geschichte*). Andere Orte (wie u.a. *Göringstadt, Warenhaus, zuhause, Eisenlager*) wurden hingegen maschinell nicht als Ort erkannt;
- mit der Geolokalisierung wurden folgende acht Orte identifiziert: *Niederösterreich, Ungarn, Polen, Wien, Salzburg, Israel, Österreich, Dachau* (vgl. Abb. 8).

EINIGE UNGENAUIGKEITEN ODER FEHLER DER AUTOMATISCHEN ANNOTATION

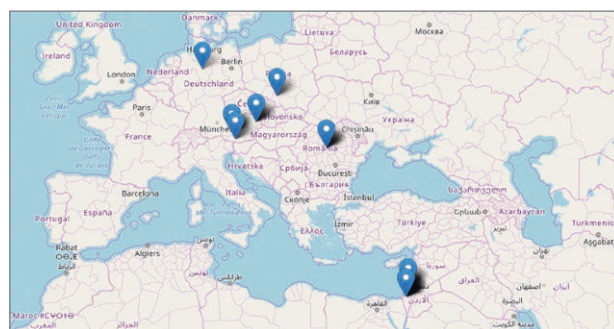


Abb. 8: Erstellung einer Karte auf der Basis der Geolokalisierung

Die aus den Geo-Daten erstellte Karte zeigt, wie wichtig diese automatisch erkannten Daten sind. Hervorzuheben ist jedoch, dass mehrere Orte fehlen (vgl. unten) und dass in einigen Fällen Orte benannt werden, die nicht mit dem Leben des/der Interviewten verbunden sind. Derzeit kann nur eine manuelle Kontextanalyse diese Fälle aufdecken.

- (3) Auf der Basis des NER-Systems wurden folgende Lexeme als Orte (B-LOC), als Institutionen/Organisationen (B-ORG), als Personen (B-PER) und als Sonstiges (B-MISC) annotiert (vgl. Tab. 2).

Die automatische Annotation hat Sondermarkierungen wie LACHEN wegen der Großschreibung als Ort markiert, sowie Personen, wie z.B. *Göring, Dollfuß, Starhemberg, Weizmann* als Orte klassifiziert und Orte wie *Israel* oder Gegenstände wie *Kippa* als Personen.

GEO-DATEN FÜR DIE ERSTELLUNG EINER KARTE

- (4) Mit automatisch erstellten Wikipedia-Links wurden sowohl Länder (*Österreich, Palästina, Israel, die Schweiz, Nordamerika, Polen, Deutschland, Libanon*), Städte (*Neunkirchen, Sopron*), Orte einer Stadt/Region (*KZ-Dachau, Schloss Schönbrunn*), als auch Personen (*Adolf Hitler, Hermann Göring, Engelbert Dollfuß, Kurt Schuschnigg*) und Organisationen (*Geheime Staatspolizei, Schutzstaffel, Betar*) identifiziert.

Dabei ist jedoch zu bemerken, dass nicht immer die gleichen Wiki-Links zu einem homonymen Ort aufgerufen werden, denn z.B. bei *Neunkirchen* wurde drei Mal der Bezirk Neunkirchen in Niederösterreich, 13 Mal der Landkreis Neunkirchen im Saarland in Deutschland und einmal die Gemeinde Neunkirchen in Nordrhein-Westfalen angegeben. Dies ist der probabilistischen, kontext-abhängigen Natur des *Named Entity Linking*-Werkzeugs geschuldet.

BEHEBUNG DER UNGENAUIGKEITEN DURCH DIE MANUELLE ANNOTATION

Die hier hervorgehobenen Ungenauigkeiten, die exemplarisch an einem Abschnitt gezeigt worden sind, die man aber im ganzen Korpus wiederfinden kann, konnten nur mit der Hilfe

der manuellen Annotation identifiziert werden. Zusätzlich wurden auch weitere Probleme bemerkt, wie die fehlende Markierung einiger Orte (19% der Orte wurde nicht erkannt) und die fehlende Unterscheidung von Bedeutungsfacetten, die man nur aus einer qualitativen Analyse der Kontexte erschließen kann.⁹

Beispielhaft möchten wir nur den Fall von ‚Schule‘ besprechen, da alle 343 Okkurrenzen des Lexems im Korpus ISW als ‚Ort‘ gekennzeichnet wurden, jedoch das Wort mit folgenden Bedeutungsfacetten im Korpus vorkommt (Kursivhervorhebung von uns):

- als ‚Lehranstalt‘, im Sinne von Institution, wie *Hauptschule* in (1), *Elementarschule* und *Landwirtschaftsschule* in (2);
- als ‚Schulgebäude‘ d.h. als metonymisch verstandener Ort in (1);
- als ‚Ausbildung, durch die jemandes Fähigkeiten auf einem bestimmten Gebiet zu voller Entfaltung kommen‘ in (3);
- als ‚Gesamtheit der Lehrer- und der Schülerschaft einer Schule‘ in (4);
- als ‚der in der Schule erteilte Unterricht‘ in (5).

- (1) Also ich bin wirklich nur in die *Hauptschule* gegangen, und zwar in der * in der * Schulgasse. Die *Schule* besteht heute noch, *3* Sie kennen sich in Wien nicht aus. Sie kennen sich in Wien aus? (ISW_E_00001)
- (2) Der Leiter des Heims war ein gewisser Doktor Lehmann, eine sehr bekannte Figur, ein deutscher Jude, * der in den zwanziger Jahren nach Palästina kam, das * Dorf gründete, es war zum Teil landwirtschaftlich, auch *Schule*, also *Elementarschule* und dann * weiter und dann noch für die älteren, jungen Leute, * dann die *Landwirtschaftsschule*, also man konnte dann weiter. (ISW_E_00004)
- (3) Hier kennengelernt, während des Krieges. Das ist eine gute *Schule* für die Ehe, man lernt sich kennen. * Ja im Leben, die schrecklichsten Situationen, ja. (ISW_E_00003)
- (4) Das war kein Studium, * aber es war eine sehr gute *Schule*. ** Und wenn ich schlechte Noten bekommen habe, hab ich mir gedacht, die *Schule* ist schlecht. Es war nicht die *Schule*, die schlecht war, sondern ich war schlecht. Die *Schule* war sehr, sehr gut. (ISW_E_00006).

- (5) Nach dem Einmarsch sind / war keine *Schule* gewesen. Für drei Wochen war keine Schule gewesen. Aber die Christen hat man ja zusammengerufen, man hat ihnen genau erklärt, wie sich zu benehmen. (ISW_E_00016).

DER EXEMPLARISCHE FALL VON ‚SCHULE‘

Für unser Projekt ist jedoch nur die Bedeutung von Schule im Sinne von ‚Schulgebäude‘, d. h. als metonymisch verstandener Ort, relevant.

Fazit und Ausblick

Zum Zweck der Auffindung von Ortsnennungen, im Sinne sowohl von Ortsnamen (Länder-, Regionen- und Städtenamen) als auch von durch Gattungswörter bezeichneten sozialen Räumen oder Teilen eines Ortes, haben wir sowohl automatische als auch manuelle Annotationsverfahren benutzt. Das Experimentieren mit *out-of-the-box*-Ansätzen, wie u. a. die Anlehnung an das Wortfeld ‚Ort‘ in GermaNet, die Herausfilterung von Geo-Koordinaten, die Markierung der Lexeme nach den Kategorien des NER-Systems und die automatische Auffindung von Wiki-Links haben viele Orte identifiziert. Die manuelle Analyse, die von zwei Experten-Annotatoren durchgeführt wurde und sich auf die Ergebnisse der automatischen Analyse stützte, hat einerseits die Unumgänglichkeit des automatischen Ansatzes betont, andererseits einige Problemfälle hervorgehoben. Dazu gehörten z.B. die partielle Auffindung von Orten, die fehlerhafte Markierung von Lexemen in mehr oder weniger festen Wortverbindungen, die Unmöglichkeit, Bedeutungsfacetten zu unterscheiden. Die Unzulänglichkeiten der automatischen Analyse konnten jedoch durch die qualitative Kontextanalyse behoben werden.

VOLLSTÄNDIGE ANNOTATION DER ORTE IM KORPUS: ERSTER SCHRITT AUF DEM WEG EINER KARTOGRAPHIE

Alle Orte aus den transkribierten Interviews des Korpus ISW sind nun aufgefunden und annotiert worden. Im nächsten Schritt streben wir die Umsetzung der entnommenen Informationen auf einer Karte an, die dann öffentlich zugänglich sein soll. ■

Literatur

- Akbik, Alan / Bergmann, Tanja / Blythe, Duncan / Rasul, Kashif / Schweiter, Stefan / Vollgraf, Roland (2019): FLAIR: an easy-to-use framework for state-of-the-art NLP. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). Minneapolis, Minnesota: Association for Computational Linguistics, S. 54-59.
- Becker, Johannes (2019): Orte und Verortungen als raumsoziologische Perspektive zur Analyse von Lebensgeschichten. In: Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 20, 1, Art. 12. <www.qualitative-research.net/index.php/fqs/article/view/3029/4338> (Stand: 23.12.2020)>.
- Betten, Anne (Hg.) (1995): Sprachbewahrung nach der Emigration – Das Deutsch der 20er Jahre in Israel. Teil I: Transkripte und Tondokumente. 1 CD. (= Phonai 42). Tübingen: Niemeyer.
- Betten, Anne (1996): Das Deutsch der 20er Jahre in Israel. In: SPRACHREPORT 4 / 1996, S. 5-10.
- Betten, Anne (2007): Zwischen Individualisierung und Generalisierung: Zur Konstruktion der Person in autobiografischen Emigranteninterviews. In: Behr, Irmtraud / Larrory, Anne / Samson, Gunhild (Hg.): Der Ausdruck der Person im Deutschen. (= Eurogermanistik 24). Tübingen: Stauffenburg, S. 173-186.
- Betten, Anne (2013): Sprachbiographien deutscher Emigranten. Die „Jeckes“ in Israel zwischen Verlust und Rekonstruktion ihrer kulturellen Identität. In: Deppermann, Arnulf (Hg.): Das Deutsch der Migranten. (= Jahrbuch des Instituts für Deutsche Sprache 2012). Berlin / Boston: de Gruyter, S. 145-191.
- Betten, Anne / Du-nour, Miryam (Hg.) (2000): Sprachbewahrung nach der Emigration – Das Deutsch der 20er Jahre in Israel. Teil II: Analysen und Dokumente. 1 CD. (= Phonai 45). Tübingen: Niemeyer.
- Betten, Anne / Du-nour, Miryam (Hg.) (2004): Wir sind die Letzten. Fragt uns aus. Gespräche mit den Emigranten der dreißiger Jahre in Israel. Unveränd. Neuaufl. der Ausg. Gerlingen, Bleicher, 1995. Gießen: Haland und Wirth.
- Brambilla, Marina / Flinz, Carolina (2019): Orte und entgegengesetzte Emotionen (Liebe und Hass) in einem Korpus biographischer Interviews (Emigrantendeutsch in Israel – Wiener in Jerusalem). In: Studi Germanici 15 / 16, S. 165-187.
- Casey, Edward S. (1993): On the phenomenology of remembering: the neglected case of place memory. In: Burton, Robert G. (Hg.): Natural and artificial minds. Albany: SUNY Press, S. 165-186.
- Deppermann, Arnulf / Schmidt, Thomas (2014): Gesprächsdatenbanken als methodisches Instrument der Interaktionalen Linguistik – Eine exemplarische Untersuchung auf Basis des Korpus FOLK in der Datenbank für Gesprochenes Deutsch (DGD2). In: Mitteilungen des Deutschen Germanistenverbandes 61, 1, S. 4-17.

- Flinz, Carolina (2019): Multiword units and n-grams naming FEAR in the Israel-Corpus. In: Corpas Pastor, Gloria/Mitkov, Ruslan (Hg.): Computational and corpus-based phraseology. Third international conference, EuroPhras 2019, Malaga, Spain, September 25–27, 2019. Proceedings. (= Lecture Notes in Computer Science 11755). Cham: Springer, S. 86–98.
- Flinz, Carolina/Moroni, Manuela Caterina (2020): Die Verwendung von „ganz“ bei der Thematisierung von Emotionen im Korpus „Emigrantendeutsch in Israel: Wiener in Jerusalem“. In: RiCOGNIZIONI, Rivista di Lingue, Letterature e Culture Moderne 7, 13, S. 105–127. <www.ojs.unito.it/index.php/ricognizioni/article/view/4472> (Stand: 23.12.2020).
- Hamp, Birgit/Feldweg, Helmut (1997): GermaNet – a lexical-semantic net for German. In: Automatic information extraction and building of lexical semantic resources for NLP applications. <www.aclweb.org/anthology/W97-0802.pdf>
- Leonardi, Simona/Thüne, Eva-Maria/Betten, Anne (Hg.) (2016): Emotionsausdruck und Erzählstrategien in narrativen Interviews. Analysen zu Gesprächsaufnahmen mit jüdischen Emigranten. Würzburg: Königshausen & Neumann.
- Leonardi, Simona/Costa, Marcella/Koesters Gensini, Sabine E./Schettino, Valentina (Hg.) (i. Ersch.): Orte und Erinnerungen: eine Kartographie des Israelkorpus. Roma: Istituto Italiano di Studi Germanici.
- Ruppenhofer, Josef/Rehbein, Iris/Flinz, Carolina (2020): Fine-grained named entity annotations for German biographic interviews. In: Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), Marseille, 11–16 May 2020. Paris: ELRA, S. 4605–4614. <www.aclweb.org/anthology/2020.lrec-1.566/> (Stand: 23.12.2020).
- Schwitalla, Johannes (2011): Narrative Formen von Fluchterzählungen deutschsprachiger emigrierter Juden in der Nazizeit. In: Thüne, Eva-Maria/Betten, Anne (Hg.): Sprache und Migration. Linguistische Fallstudien. (= Lavori interculturali sul tedesco 4). Rom: Aracne, S. 17–51.
- Schwitalla, Johannes (i. Ersch.): Erfahrungsräume in prekären Situationen. In: Leonardi/Costa/Koesters Gensini/Schettino (Hg.), o.S.
- Thüne, Eva-Maria (2009): Dinge als Gefährten. Objekte und Erinnerungsgegenstände in Bettens Israel-Korpus der ersten Generation. In: Dannerer, Monika/Mausser, Peter/Scheutz, Hannes/Weiss, Andreas (Hg.): Gesprochen – geschrieben – gedichtet. Variation und Transformation von Sprache. Festschrift für Anne Betten zum 65. Geburtstag. (= Philologische Studien und Quellen 218). Berlin: Erich-Schmidt, S. 189–204.
- Thüne, Eva-Maria/Leonardi, Simona (2011): Wurzeln, Schnitte, Webe-muster. Textuelles Emotionspotenzial von Erzählmetaphern am Beispiel von Anne Bettens Interviewkorpus „Emigrantendeutsch in Israel“. In: Kohlross, Christian/Mittelmann, Hanni (Hg.): Auf den Spuren der Schrift. Israelische Perspektiven einer internationalen Germanistik. (= Conditio Judaica 80). Berlin/Boston: de Gruyter, S. 229–246.

Internetquellen

- Datenbank für Gesprochenes Deutsch <<https://dgd.ids-mannheim.de>>.
Orte und Erinnerungen. Eine Kartographie des Israelkorpus. <<https://kartografiedesisraelkorpus.wordpress.com>>.
Zotero <www.zotero.org/groups/2219390/israelkorpus/library>.

Anmerkungen

- 1 Korpusname (Korpuskennung): Emigrantendeutsch in Israel (IS), Emigrantendeutsch in Israel: Wiener in Jerusalem (ISW) und Zweite Generation deutschsprachiger Migranten in Israel (ISZ).
- 2 Vgl. dazu die Beschreibung der Korpora in der DGD. Zu früheren Berichten Bettens über das Forschungsprojekt vgl. u. a. Betten (1996, 2013). Zu einer laufend aktualisierten Gesamtbibliographie zu den Israelkorpora mit Nachweis der jeweils analysierten Interviews siehe <www.zotero.org/groups/2219390/israelkorpus/library>.
- 3 Seit unserer Studie zur Erkennung von Ortsnennungen, deren Ergebnisse wir auch auf der DHD-Tagung 2020 (vgl. <<https://dhd2020.de>>, Stand: 23.12.2020) vorgestellt haben, haben wir die Daten auch mit einem feinkörnigen Schema von *Named Entities* annotiert (Ruppenhofer/Rehbein/Flinz 2020).
- 4 Oben: Shaul Otto und Hanna Baumann, Shoshana Beer, Paul Rudolf Beer. Unten: Jeannette und Erich Goldstein, Ari Rath.
- 5 Vgl. <www.sfs.uni-tuebingen.de/GermaNet/>. (Stand: 23.12.2020).
- 6 Erläuterungen zu den Abkürzungen: GN steht für *GermaNet*; GEO für Geo-Koordinaten; PER für Person; LOC für Orte, ORG für Organisationen, MISC für unterschiedliche Informationen, TIME für Zeitangaben, NER für *Named Entity Recognition System*, NEL für *Named Entity Linking*.
- 7 Die Kappa-Werte für diese Transkripte sind zwischen 0,748 und 0,793.
- 8 Interview mit Naftali Jaul, geb. 1915 in Neunkirchen (Niederösterreich) als Anton Jaul.
- 9 Ein automatisches Werkzeug zur Wortbedeutungsdisambiguierung mit ausreichender Abdeckung und Akkuratheit für Ortsnennungen steht z. Zt. nicht zur Verfügung.

Bildnachweise

- Abb. 2: Betten/Du-nour (Hg.) (2004)
Abb. 3: <<https://kartografiedesisraelkorpus.wordpress.com>>
Abb. 5: Anne Betten
Abb. 8: Josef Ruppenhofer ■