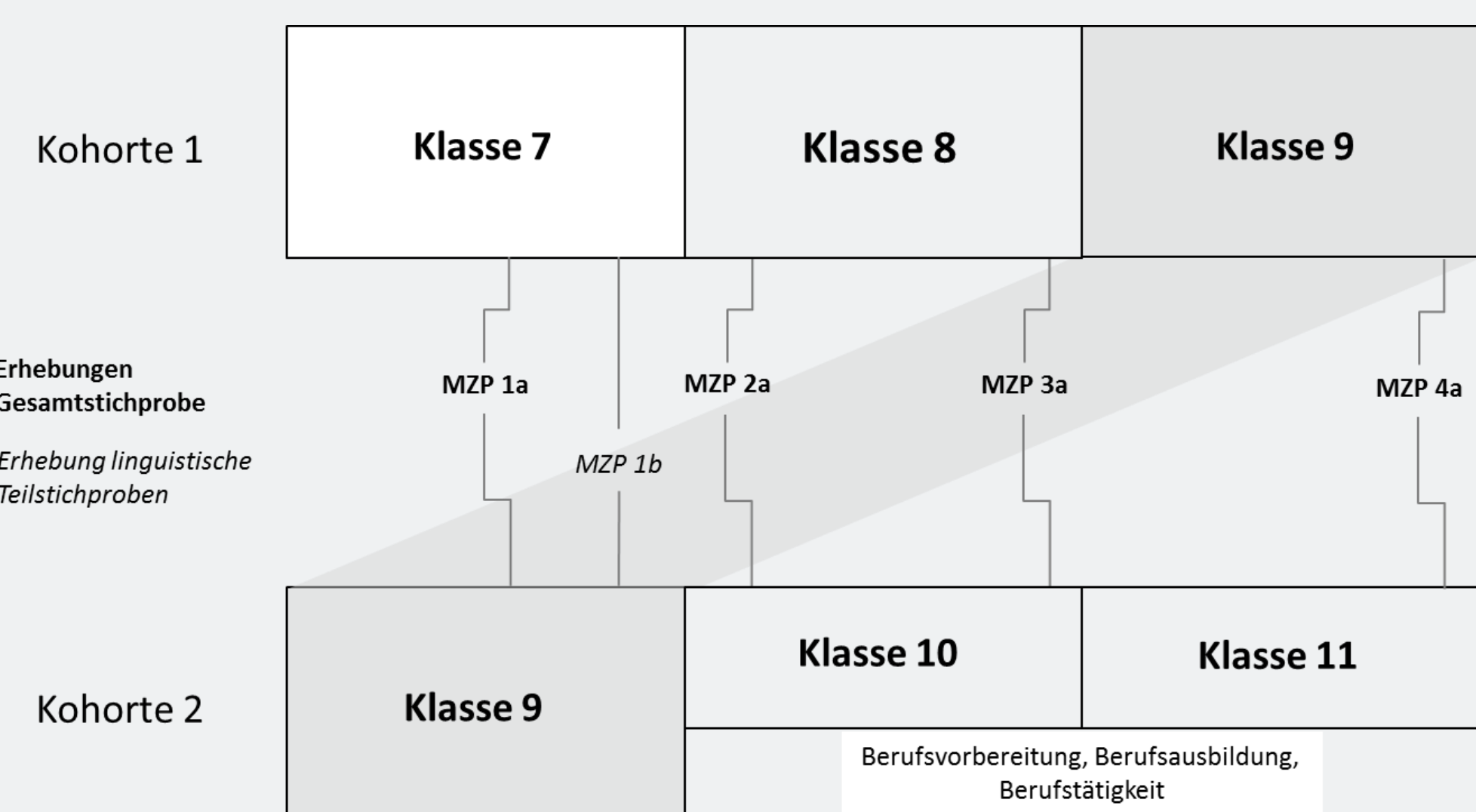


SIND WIR INTERESSANT FÜR KI – ODER: IST KI INTERESSANT FÜR UNS?

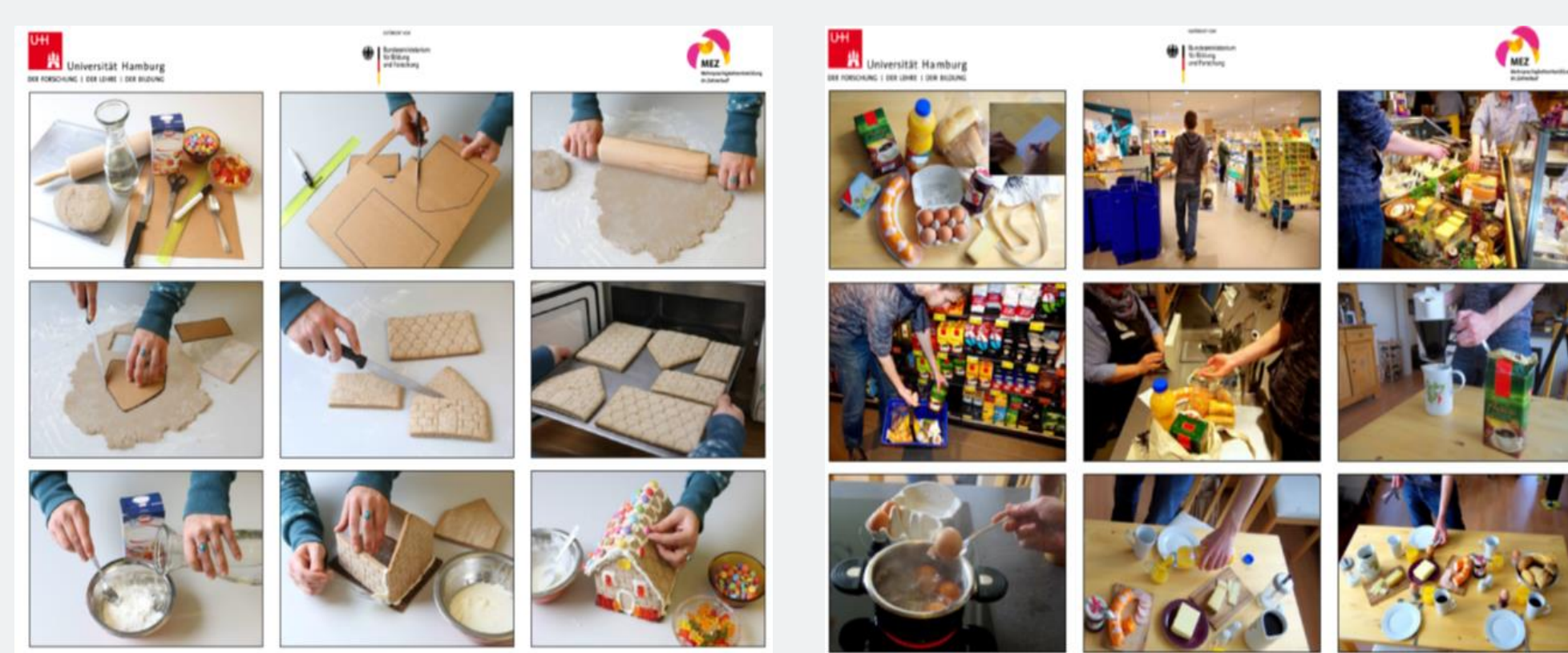
ANALYSEN DES TEXTKORPUS „MEHRSPRACHIGKEITSENTWICKLUNG IM ZEITVERLAUF (MEZ)“

DAS PROJEKT:

Die MEZ-Studie untersucht die Entwicklung mehrsprachiger Literalität bei Schülerinnen und Schülern in der Sekundarstufe¹⁾. Die Jugendlichen kommen aus deutsch-russischen, deutsch-türkischen oder einsprachig deutschen Familien. Sie lernten Englisch als erste Fremdsprache, gelegentlich auch Französisch oder Russisch als zweite Fremdsprache.

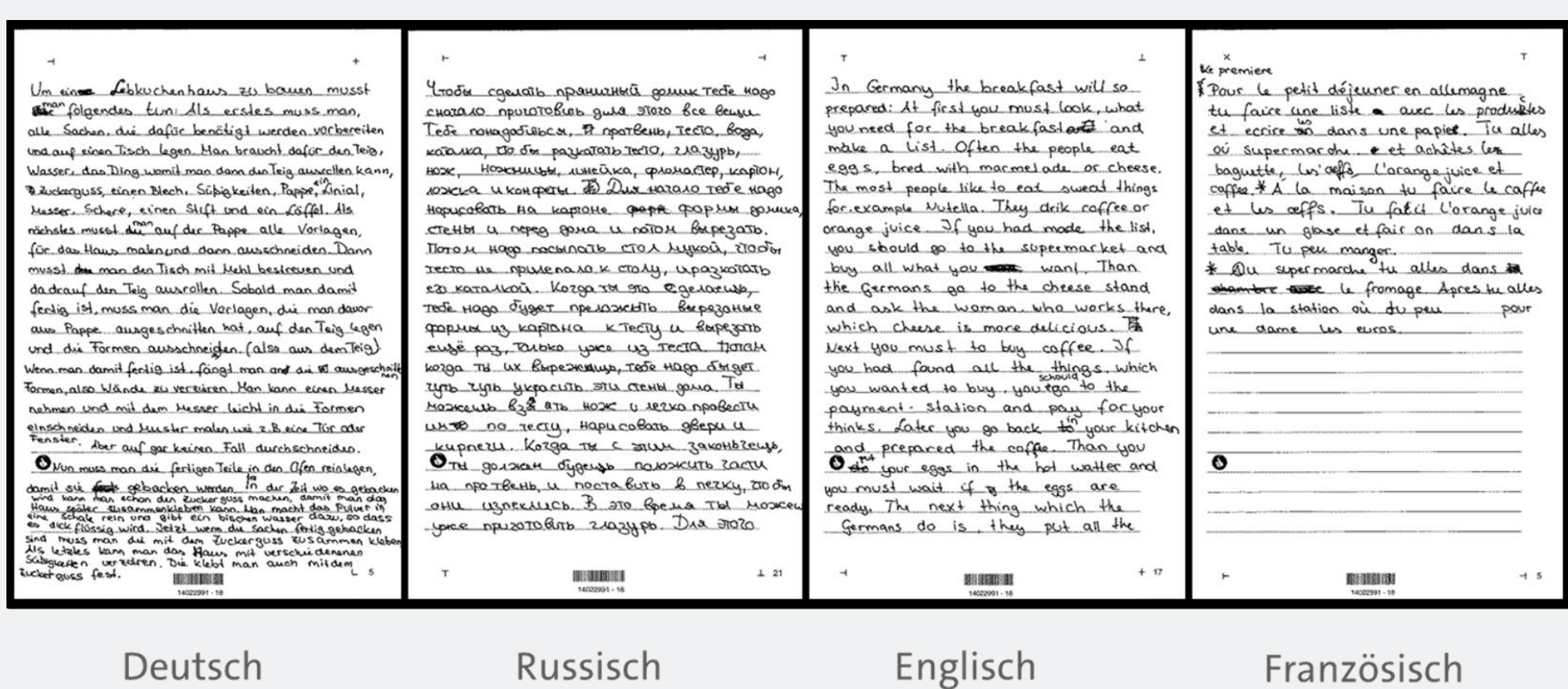


Die Anfangsstichprobe umfasste 2103 Proband(inn)en, deren schriftsprachliche Fähigkeiten in vier Messzeitpunkten mit Sprachtests gemessen wurden. MEZ bewertete die Schriftfähigkeiten in den Familiensprachen Deutsch, Russisch und Türkisch; in Englisch als 1. Fremdsprache sowie bei einigen Französisch oder Russisch als 2. Fremdsprache.



Beispiele für die verwendeten Schreibimpulse

Die Schreibkompetenzen wurden mit der „MEZ-Schreibaufgabe für Jugendliche“ erhoben. Jede Parallelversion verwendet eine Serie von neun Fotos als visuelle Stimuli und je eine darauf bezogene Aufgabenstellung. Ein Schreib-Beispiel einer Schülerin:



UNSER PROBLEM:

Die Auswertung des komplexen Korpus ist sehr aufwendig – u.a., weil intensive Qualitätskontrollen erforderlich sind. Das Verfahren wäre aber weltweit einsetzbar, da es den Qualitätsanforderungen an Diagnostik einer mehrsprachigen Schülerschaft standhält – also einer Lehr- und Lernkonstellation, die in den meisten Ländern der Welt besteht.

Das Korpus enthält 16.859 handgeschriebene Texte als Bilddateien und Transkriptionen, die die handschriftliche Schreibweise – also oft: originelle Orthographie – genau wiedergeben.

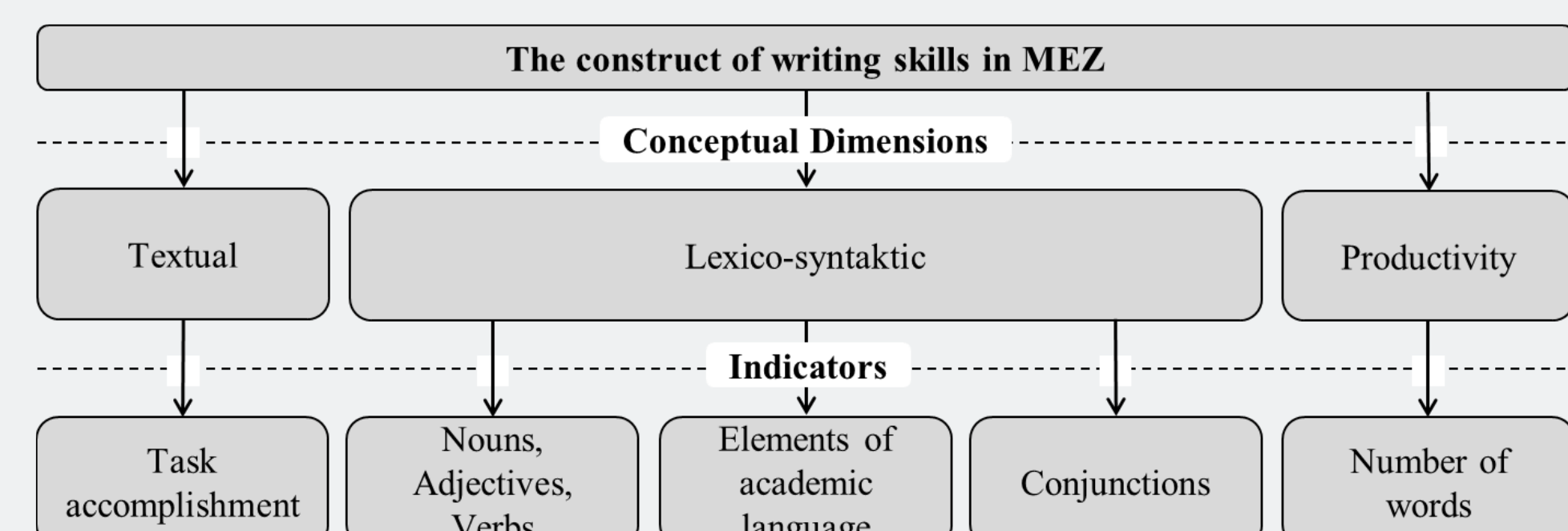
Hinweis	Originaltext
Nicht auf Zeilenumbruch des Ursprungstexts achten, aber formale Gliederung des Originals beibehalten (Absatz, tabellarische Darstellung, Auflistung, etc.)	<i>On pic-nique avec son copain. On avec le (essen) On avec le (bonbonnet) de la</i>
Rechtschreibung und Zeichensetzung des Originaltextes STRIKT beibehalten	<i>- Apples, - Tomatos - Bananas - and soft.</i>
nicht erkannte Wörter: Tag in eckiger Klammer; durchgehend geschrieben: [?nomen], [?verb], [?adjektiv] [?unbekannt]	<i>The things that we are buy</i>
Teil des Wortes/Wort nicht erkannt: [?]	<i>prepa id.</i>
Silbentrennung: [-]	<i>eine Hob platte einm Akk... eine Nück- sage</i>
durchgestrichene Zeichen/Wörter	<i>On avec</i>
Wort/Zeichen über der Zeile	<i>you go to for on Picnic have after eh fromage</i>

Der MEZ-Bestand von Schülertexten und angefertigten Transkripten (Mittelwerte und Standardabweichungen der Textlänge [Anzahl der Wörter] in Klammern):

	Deutsch	Herkunfts- sprache Russisch	Herkunfts- sprache Türkisch	Fremd- sprache Englisch	Fremd- sprache Französisch	Fremd- sprache Russisch	Gesamt	
Welle 1	Auswertbare Texte (195;57)	248 (98;50)	421 (98;40)	1.741 (137;59)	661 (82;48)	58 (91;43)	4.880	
	Transkripte	1.770	288	441	1.770	693	61	5.023
Welle 2	Auswertbare Texte (166;42)	235 (92;40)	469 (86;34)	1.680 (130;50)	652 (81;47)	60 (85;39)	4.812	
	Transkripte	1.718	276	489	1.716	681	60	4.940
Welle 3	Auswertbare Texte (148;47)	182 (73;29)	398 (76;34)	1.584 (153;58)	558 (90;50)	55 (64;27)	4.368	
	Transkripte	1.607	246	436	1.620	578	55	4.542
Welle 4	Auswertbare Texte (162;56)	123 (93;30)	243 (95;32)	1.048 (158;57)	269 (109;51)	33 (92;26)	2.799	
	Transkripte	1.095	153	275	1.091	324	34	2.972
Gesamt	Auswertbare Texte	6.141	788	1.531	6.053	2.140	206	16.859
	Transkripte	6.190	963	1.641	6.197	2.276	210	17.477

UNSERE FRAGEN:

In den Analysen nutzen wir ein theoretisches Modell der Schreibleistung in mehreren Sprachen zur Textqualitätsbewertung. Dies ist für die verschiedenen Sprachen identisch – erlaubt also auch vergleichende Aussagen über die Qualität der Schreibprodukte. Ist das eine interessante Anwendungsmöglichkeit für KI?



Uns interessiert der Anwendungsaspekt in zwei Richtungen:

- 1) Sind die MEZ-Daten interessant für das Trainieren von KI?
 - Unsere Texte wurden auf Basis eines theoretisch abgesicherten Modells bewertet. Die Ergebnisse können als Evaluationskriterium für den Trainingserfolg der KI genutzt werden.
- 2) Welchen Nutzen könnte die KI für die Automatisierung unserer Auswertungen haben?
 - Ist es möglich unsere Texte mit einem KI-Modell auszuwerten, das den Regeln unseres theoretischen Kompetenzmodells folgt?
 - Kann KI selbständig zusätzliche Strukturen in den Texten entdecken, die deren Qualität anzeigen?



Wir sind gespannt...

Bei Interesse, Fragen und Rückmeldungen schreiben Sie uns: ingrid.gogolin@uni-hamburg.de

INGRID GOGOLIN
NORA DÜNKEL
THORSTEN KLINGER
KSENIJA PERSHINA
BIRGER SCHNOOR
ANOUK TICHELOVEN
IRINA USANOVA

¹⁾ MEZ wurde 2014 bis 2019 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert. Im Anschluss erfolgte eine Förderung durch das BMBF für das Folgeprojekt *Mehrsprachigkeit an der Schwelle zum Beruf (MEZ-2)* sowie für das Forschungsprogramm *Multiliteralität als Arbeitsmarktressource (MARE)*