

Emmerich Kelih, Universität Wien, Österreich

„Objektive“ und „subjektive“ Frequenz von Internationalismen im Kroatischen

The article deals with the frequency of internationalisms in Croatian from a structural linguistics perspective, taking into account usage-based and psycholinguistic approaches (cf. Backus, 2013 for an overview). Specifically, two different kinds of frequency will be distinguished: The first one is the ‘objective’ frequency, which is taken from a Croatian frequency dictionary (cf. Moguš et al., 1999). The second one refers to the speaker-based, ‘subjective’ assessments of the frequency, which is taken from the Croatian psycholinguistic database (cf. Peti-Stantić et al., 2021). In addition, also the polytextuality (dispersion), i.e. the occurrence of the internationalisms in different functional style, will be discussed as reliable criterion for determination of a core inventory of Croatian internationalisms. Finally, some correlations between the subjective assessment and the polytextuality will be emphasized. The combination of them helps in the determination of the high-frequency inventory of Croatian internationalism, which in turn are a reliable indicator in how far the Croatian core vocabulary is influenced lexically because of language contact respectively puristic attitudes. This inventory of internationalisms is not only of interest from a linguistic point of view but also provides a solid basis for foreign language didactics. It can be assumed that internationalisms can facilitate text comprehension and therefore play a special role in language acquisition. In addition, it can be shown in how much objective and subjective frequency and polytextuality are important selection criteria when creating basic vocabulary.

Keywords: objective frequency, subjective frequency, internationalism, frequency dictionary, polytextuality, Croatian

Der Artikel befasst sich mit der Häufigkeit von Internationalismen im Kroatischen aus einer systemlinguistischen Perspektive, unter Berücksichtigung von gebrauchorientierten und psycholinguistischen Ansätzen (vgl. Backus, 2013 für einen Überblick). Konkret werden zwei verschiedene Arten von Häufigkeiten unterschieden: Die erste ist die „objektive“ Frequenz, die einem kroatischen Frequenzwörterbuch entnommen wird (vgl. Moguš et al., 1999). Die zweite bezieht sich auf die sprecherbasierte, „subjektive“ Einschätzung der Frequenz, die der kroatischen psycholinguistischen Datenbank entstammt (vgl. Peti-Stantić et al., 2021). Darüber hinaus wird auch die Polytextie (Dispersion), d.h. das Vorkommen der Internationalismen in verschiedenen Funktionalstilen als zuverlässiges Kriterium für die Bestimmung des Kerninventars kroatischer Internationalismen diskutiert. Abschließend werden vorhandene Korrelationen zwischen der subjektiven Frequenz und der Polytextie aufgezeigt. Die Kombination dieser Faktoren hilft bei der Bestimmung des hochfrequenten Inventars kroatischer Internationalismen, das wiederum ein zuverlässiger Indikator dafür ist, inwieweit der kroatische Kernwortschatz durch den Sprachkontakt und puristische Attitüden lexikalisch beeinflusst ist. Dieses Inventar an Internationalismen ist nicht nur aus sprachwissenschaftlicher Sicht von Interesse, sondern bietet auch eine solide Grundlage für die Fremdsprachendidaktik. Es ist davon auszugehen, dass Internationalismen das Textverständnis erleichtern können und im Spracherwerb daher eine besondere Rolle spielen. Darüber hinaus sind objektive, subjektive Frequenz und Polytextie wichtige Selektionskriterien bei der Erstellung von Grund- bzw. Basiswortschätzen.

Schlüsselwörter: objektive Frequenz, subjektive Frequenz, Internationalismen, Häufigkeitswörterbuch, Polytextie, Kroatisch



1. Einleitung: „Objektive“ vs. „subjektive“ Frequenz

In der Fremdsprachendidaktik gibt es seit langem eine intensive Diskussion über die Bedeutung der Vorkommenshäufigkeit von lexikalischen Einheiten (Wortfrequenz) im weitesten Sinne.¹ Besondere Bedeutung erlangt diese Frage in der Diskussion um Selektionskriterien bei der Erstellung von Basis- bzw. Grundwortschätzen einer Sprache. Der Fokus des vorliegenden Beitrags liegt allerdings weniger auf der Frage der Selektion, sondern der Ausgangspunkt für die folgende Abhandlung ist eine vor allem in der Psycholinguistik getroffene Unterscheidung zwischen „objektiver“ und „subjektiver“ Frequenz, die in Kap. 1.1 und 1.2 in groben Zügen vorgestellt und diskutiert wird. Im Folgenden werden diese beiden Frequenzarten am Beispiel des Kroatischen ausführlich diskutiert. Vorab erfolgt jedoch eine Einschränkung auf einen bestimmten Teilbereich der Lexik, nämlich den Lehnwortschatz, und zwar auf die Internationalismen. Diese spielen beim Fremdspracherwerb eine besondere Rolle, insbesondere im Hinblick auf die Perzeption und das Verstehen. Internationalismen sind vor allem am Beginn des Spracherwerbs von Interesse, da mit deren Hilfe die entsprechenden zielsprachlichen Synonyme eingeführt werden können. Generell sind sie aber auch ein (grober) Indikator für den „Internationalisierungsgrad“ einer Sprache, d.h. den Einfluss von Sprachkontakten auf die Lexik. Insgesamt verbindet der vorliegende Aufsatz somit ein genuin linguistisch-theoretisches Interesse im Sinne eines gebrauchsbasierten Ansatzes mit fremdsprachendidaktischen Anliegen.

1.1 Zur theoretischen Bedeutung von „objektiver“ lexikalischer Frequenz

Im Zusammenhang mit der „objektiven“ Wortfrequenz geht es oft um konkrete Fragen der empirischen Bestimmung und allgemein um angewandte Aspekte (z.B. Dechiffrierung, Sprachtechnologie). Eine weitverbreitete Art der Bestimmung der „objektiven“ Frequenz bezieht sich auf text- oder korpusbasierte Analysen. Konkrete Ergebnisse in diesem Zusammenhang sind u.a. Frequenzwörterbücher, die auch bei der Erstellung von Basis- und Grundwortschätzen eine Rolle spielen.

Bevor wir jedoch in Kapitel 1.2 auf diesen Aspekt näher eingehen, sollen einige grundlegende Annahmen zur Worthäufigkeit genannt werden, die sich aus den Erfahrungen der gebrauchsbasierten Linguistik, der Psycho- und Kognitionslinguistik, aber auch der quantitativen Linguistik ableiten lassen. Aus einer Vielzahl von Studien ist seit langem bekannt (vgl. Köhler, 2005a; Tuldava, 1998), dass die hochfrequente Lexik einer Sprache durch ganz spezifische Merkmale gekennzeichnet ist und dass eine nach Frequenzen geordnete Menge von Lexemen aus theoretischer Sicht durchaus von Interesse ist.

Ermittelt man z.B. die Häufigkeit der in einem Text vorkommenden Wörter, so stellt man fest, dass vor allem die Synsemantika sehr häufig sind. Ihre Funktion besteht offensichtlich vor allem darin, die morphosyntaktische Struktur des Textes zu organisieren, die sich durch eine hohe Zahl von Wiederholungen auszeichnet (z.B. aufgrund der Verwendung von bestimmten hochgradig

¹ Der Autor bedankt sich herzlich bei zwei anonymen Reviewern für die Vielzahl von konstruktiven Kommentaren und kritischen Bemerkungen, die zu einer nachhaltigen Verbesserung des Artikels beigetragen haben. Die eruierten Rohdaten (Inventar an kroatischen Internationalismen) sind als pdf-Datei unter <https://phaidra.univie.ac.at/o:2138469> veröffentlicht.

polyfunktionalen Verben, insbesondere aber Hilfsverben, Pronomen, Präpositionen, Konjunktionen usw.). Aus einem Anwendungsbedürfnis entsteht die Notwendigkeit nach Wiederholung, die sich in unterschiedlichen Formen der Häufigkeit äußert. Ein weiteres zentrales Merkmal natürlichsprachlicher Texte ist die Redundanz, die notwendig ist, um Übertragungssicherheit in der Kommunikation zu gewährleisten. Diese wird ebenfalls durch die Wiederholung von Elementen erreicht.

Aber nicht nur Synsemantika unterliegen einem Wiederholungsdruck, sondern aus vielen Untersuchungen zum so genannten Zipfschen Gesetz² (vgl. dazu den informativen Überblick in Pustet, 2004) ist bekannt, dass das Vorkommen von Autosemantika³ in Texten einem bestimmten statistischen Muster folgt: Einige wenige Wörter kommen in einem Text sehr häufig vor, die meisten dagegen sehr selten, viele davon sogar nur als *Hapax legomena* (Wörter, die nur einmal im Text bzw., je nach Definition, in einer Sprache nur einmal vorkommen). Dieses mathematisch exakt modellierbare Zusammenspiel von hoher und niedriger Frequenz ist konstitutiv für natürlichsprachliche Texte (vgl. zusammenfassend Köhler, 2005b).

Das Zipfsche Gesetz enthält aber nicht nur Aussagen über den Zusammenhang von Rang und Häufigkeit und die spezifische nichtlineare Form der Verteilungskurve, sondern liefert auch Informationen über die Häufigkeit und Länge/Kürze sprachlicher Ausdrücke. Betrachtet man in einem Text z.B. nur Autosemantika, so fällt auf, dass diese sich durch ihre relative Kürze (Wortlänge) auszeichnen. Das heißt, je häufiger ein (autosemantisches) Wort in einem Text (bzw. in einer Sprache) vorkommt, desto kürzer ist es. Für seltene Wortformen gilt, dass sie in der Regel länger sind. Dieser Zusammenhang zwischen Frequenz und Länge kann durchaus im Sinne eines sprachübergreifenden Kodierungsmechanismus verstanden werden, der direkte Auswirkungen auf die Produktion, aber auch auf die Rezeption sprachlicher Einheiten hat.

Zum Abschluss dieser kurzen theoretischen Einführung soll noch ein dritter, aus unserer Sicht sehr wichtiger Aspekt erwähnt werden. Es gibt nicht nur einen Zusammenhang zwischen Häufigkeit und Länge, sondern es ist auch bekannt, dass kürzere sprachliche Einheiten in der Regel auch über mehr Bedeutungen verfügen (vgl. Levickij, 2005 mit einer Übersicht). Es besteht somit ein Zusammenhang mit der Polysemie sprachlicher Ausdrücke, die sich – kurz gesagt – dadurch erklären lässt, dass Wörter mit längeren Morphemketten einen höheren Spezifizierungsgrad aufweisen als kürzere Einheiten.

Insgesamt ist die Worthäufigkeit nicht nur ein Indikator für Gebrauch und Verwendung, sondern mit ihr sind weitreichende Fragen der Kodierung sprachlicher Einheiten und kognitiver Verarbeitungsmechanismen verbunden. Die angeführten Zusammenhänge zwischen Frequenz, Länge

² Korrekterweise muss man von den Zipfschen Gesetzen sprechen, da – wie im Fließtext dargelegt – mindestens vier unterschiedliche Merkmale (Rang, Frequenz, Länge und Polysemie) erfasst werden.

³ Hinsichtlich der Internationalismen wären auch die vorhandenen kroatischen Synonyme zu berücksichtigen bzw. es wäre zu untersuchen, inwiefern diese zur sogenannten *lexical density* beitragen. Dies ist in der Tat ein Desiderat, das es in Zukunft näher zu untersuchen gilt.

und Bedeutungskomplexität sind auch mit bekannten Konzepten, wie Entrenchment,⁴ in Verbindung zu bringen (vgl. dazu im Detail Kelih, 2023, 122–129).

1.2 Zur theoretischen Bedeutung von subjektiver Frequenz

Für die Psycholinguistik und die kognitive Linguistik scheint zu gelten, dass keine Variable so ausführlich diskutiert worden ist wie die Worthäufigkeit. Als mögliche „objektive“ Frequenzzählungen gelten, wie bereits erwähnt, Auswertungen der Häufigkeit von Wörtern in natürlichen Texten, Textstichproben und Korpora. Dabei geht es um die so genannte „objektive“ oder „durchschnittliche“ Häufigkeit eines Wortes in einer Sprache. Aus linguistischer Sicht ist die Annahme einer „durchschnittlichen“ Worthäufigkeit jedoch zweifelhaft, da zentrale Faktoren wie Stichprobengröße, Repräsentativität und Ausgewogenheit der untersuchten Texte (einem zentralen Problem der Korpuslinguistik, wie in Stefanowitsch, 2020, 22–45 ausführlich diskutiert) berücksichtigt werden müssen.

Es ist daher verständlich, dass man sich vor allem in der Psycholinguistik bemüht hat, neben den „objektiven“ Häufigkeiten auch die subjektive Einschätzung der Häufigkeit bzw. Geläufigkeit („subjective ratings of familiarity“) von Wörtern zu untersuchen. Um diese zu ermitteln, werden Versuchspersonen, grob gesprochen, Wörter mit der Bitte vorgelegt, z.B. die hohe, mittlere und niedrige Häufigkeit des Wortes auf einem vorher gewählten Skalenniveau einzuschätzen. Auch diese Form der Worthäufigkeitsmessung ist mit verschiedenen Schwierigkeiten verbunden. Unter anderem ist unklar und es wird intensiv diskutiert, inwieweit die Anzahl der Bedeutungen, die Kontextspezifität, der Grad der Abstraktheit/Konkretheit, die semantische Vernetzung, morphologische Eigenschaften usw. einen Einfluss auf die subjektiven Urteile haben (vgl. dazu u.a. Anstatt, 2016; Baayen, Feldman & Schreuder, 2006). Die subjektive Häufigkeit ist darüber hinaus klarerweise von sozialen Variablen wie Alter, Geschlecht, Schulbildung (vgl. hierzu Balota, Pilotti & Cortese, 2001), aber auch beruflicher Spezialisierung abhängig.

Es ist davon auszugehen, dass Sprecher*innen eine sehr gute Kompetenz für die subjektive Einschätzung der Häufigkeit von Wörtern haben („subjective frequency estimates“).⁵ Zudem kann, wie McGee (2006, 95–99) zeigt, angenommen werden, dass die subjektive Einschätzung von Worthäufigkeiten recht gut mit den objektiv feststellbaren Häufigkeiten in Sprachkorpora übereinstimmt, auch wenn bei bestimmten Substantiven, Synonymen, relativ bekannten Konzepten und niedrigfrequenten Einheiten Ausnahmen zu beobachten sind und sprachspezifisch zu eruieren sind.

Zusammenfassend lässt sich festhalten, dass Frequenz im linguistischen Kontext unterschiedlich interpretiert und vor allem operationalisiert werden kann. Je nach Forschungsinteresse ist zu entscheiden, ob mit Hilfe objektiver Frequenzen der sprachliche Output analysiert wird oder ob

⁴ Das ist in der Linguistik intensiv diskutierte Konzept von *entrenchment* lässt sich im Deutschen als kognitive Verfestigung übersetzen. Die kognitive Verfestigung ergibt sich aus den im Sprachgebrauch beobachtbaren und wiederkehrenden sprachlichen Mustern. Sie treten als verfestigte Einheiten auf, sind als mentale/kognitive Einheiten abgespeichert und werden in dieser Form bei Bedarf als Ganzes reproduziert (vgl. dazu u.a. Ziem & Lasch, 2013, 16).

⁵ Geläufigkeit (*familiarity*) und Wortfrequenz sind in der Psycholinguistik und kognitiven Wissenschaften fest etabliert, wenngleich auch Konzepte wie Bekanntheit, produktive bzw. rezeptive Kenntnis, Verwendungshäufigkeit, Akzeptabilität usw. eng verbunden sind.

mit Hilfe subjektiver Frequenzen eine stärker sprecher*innenzentrierte Perspektive eingenommen wird. Darüber hinaus ist zu beachten, dass beide Arten von Frequenzen in gewisser Weise korrelieren und sich zudem überlappen. Diese Dimension wird in Kapitel 2 anhand der vorhandenen Daten zu objektiven und subjektiven Frequenzen im Kroatischen empirisch demonstriert. Zuvor soll jedoch im nächsten Kapitel der Fokus der vorliegenden Untersuchung, nämlich Entlehnungen bzw. insbesondere die im Mittelpunkt stehenden Internationalismen, näher vorgestellt werden.

1.3 Häufigkeit von Entlehnungen (insbesondere Internationalismen)

Der Wortschatz einer Sprache besteht aus unterschiedlichen Schichten. In der historisch-vergleichenden Linguistik wird zwischen einem Erb- und Lehnwortschatz unterschieden. Diese (einfache) binäre Trennung wollen wir im Folgenden aufnehmen und uns aber ausschließlich auf den sogenannten Lehnwortschatz konzentrieren.

Aus den Ausführungen in Kapitel 1.2 lässt sich ableiten, dass hochfrequenten Entlehnungen in einer Sprache eine besondere Bedeutung zukommt, zumal sich auch häufige Entlehnungen hinsichtlich ihrer Länge und ihres Bedeutungspotenzials von selteneren Entlehnungen unterscheiden sollten. Darüber hinaus ist das Inventar hochfrequenter Einheiten an sich für den Fremdsprachenerwerb von Interesse. Im vorliegenden Beitrag steht jedoch weniger diese theoretische als die empirische Dimension im Vordergrund. Im Einzelnen geht es um die Frage, welchen quantitativen Anteil Internationalismen am hochfrequenten Wortschatz einer Sprache haben. Diese Frage ist jedoch kein Selbstzweck, sondern es ist zumindest in Form einer Arbeitshypothese davon auszugehen, dass im Grundwortschatz einer Sprache relativ wenige Entlehnungen zu finden sind und daher von einer relativen Entlehnungsresistenz auszugehen ist. In einer Vielzahl von Studien konnte gezeigt werden, dass lexikalische Entlehnungen in einen Grundwortschatz nur bei relativ intensivem Sprachkontakt stattfinden (vgl. Thomason, 2001, 70), während bei oberflächlichem Kontakt der lexikalische Transfer in erster Linie auf einzelne Nomen beschränkt bleibt. So konnte z.B. im Rahmen des Projekts *WOLD, The World Loanword Database* von Haspelmath & Tadmor (2009) auf der Basis von über 40 untersuchten Sprachen gezeigt werden, dass erstens der Anteil von Entlehnungen im Grundwortschatz sehr stark variiert (de facto zwischen vier und 65%) und dass es zweitens innerhalb des Grundwortschatzes bestimmte semantische Bereiche gibt, die besonders „offen“ für Entlehnungen sind (wie z.B. der Bereich Religion, vor allem aber „konkrete“ Lebensbereiche wie Kleidung, Haushalt, technische Innovationen usw.). Umgekehrt gibt es aber auch nicht wenige semantische Wortfelder, die tatsächlich eine hohe Entlehnungsimmunität aufweisen, wie z.B. im Falle von Verwandtschaftsbezeichnungen, Ausdrücken für räumliche und zeitliche Beziehungen, Funktionswörtern usw. In diesen lexikalisch-semantischen Bereichen finden sich in der Regel nur wenige oder gar keine Entlehnungen. Hervorzuheben ist, dass diese Ergebnisse auf einer (im Detail nicht unumstrittenen) Definition von Basiswortschatz beruhen, bei der es darum geht, im weitesten Sinne wichtige, aber stilistisch neutrale, alltagssprachliche, nicht kulturgebundene Wörter zu untersuchen (vgl. dazu Tadmor, Haspelmath & Bradley, 2010; Zenner, Speelman & Geeraerts, 2014). Allerdings ist zu ver-

merken, dass bei der Erstellung von Basiswortschätzen weder objektive noch subjektive Häufigkeiten zur Auswahl herangezogen werden (dies gilt auch für den im *WOLD* verwendeten Wortschatz, dessen Auswahl sich an älteren Ansätzen der vergleichend-historischen Sprachwissenschaft orientiert, aber eigentlich im Detail nicht begründet ist, vgl. dazu Kelih, 2023, 91–93). Daher sei an dieser Stelle noch einmal auf die Diskussion um einen durch Häufigkeiten gewichteten Grundwortschatz hingewiesen, der eine aus unserer Sicht sinnvolle und nachvollziehbare Möglichkeit für die Transparenz der Auswahl darstellt.

Dabei steht die Idee im Vordergrund, dass bei der Erstellung von Basis- bzw. Grundwortschätzen die entsprechende Vorkommenshäufigkeit berücksichtigt werden muss, da es auf diese Weise gelingt, eine relevante und wichtige Auswahl von Wörtern zu treffen, die für die Bewältigung grundlegender kommunikativer Bedürfnisse notwendig ist (vgl. dazu u.a. Krohn, 1992; Haderlein, 2008; Bubenhofer et al., 2015). Ausgehend von dieser Idee soll nun im Folgenden in Form einer Pilotstudie das Inventar der Internationalismen im Kroatischen in einem frequenzbasierten Wortschatz ermittelt werden. Dabei wird u.a. zu diskutieren sein, ob der ermittelte Anteil tatsächlich als Indikator für das Ausmaß des lexikalischen Einflusses von „außen“ gelten kann und damit die Intensität des Sprachkontakts widerspiegelt. Darüber hinaus wird zu demonstrieren sein, welche Bedeutung der Berücksichtigung der subjektiven Frequenz, aber auch der sog. Polytextie (= das Vorkommen in verschiedenen Funktionalstilen) zukommt.

Die Frage nach den objektiven und subjektiven Häufigkeiten von Entlehnungen im Kroatischen ist unseres Wissens noch nicht systematisch untersucht worden. In der folgenden empirischen Untersuchung wird daher eine weitere Einschränkung, nämlich die Fokussierung auf sogenannte Internationalismen (vgl. dazu Kolwa, 2003; Braun, 1990 und Schaeder, 1990 mit einem Überblick) vorgenommen.

Die verschiedenen Möglichkeiten, Internationalismus zu definieren, sind Gegenstand einer langen und vielschichtigen Diskussion. In Glück (2010, 301–302) wird Internationalismus als ein Lehnwort definiert, das in sehr vielen Sprachen gebräuchlich ist, wie z.B. *Taxi*, *Hotel*, *Computer* usw. Diese seien für die internationale Verständigung sehr nützlich, zumal sie, wie eine etwas andere, sehr allgemeine Definition von Internationalismus lautet, „[...] sprachliche Einheiten, die in Ausdruck und Inhalt (ggf. nur in Struktur und Inhalt) in mehr als einer Sprache gleich oder ähnlich vorkommen [...]“ (vgl. Bergmann, 1995, 272, zitiert nach Schaeder, 2003, 86).

Diese Definition ist jedoch im slawischen Kontext nur eingeschränkt anwendbar, da viele slawische Sprachen in ihrem Erbwortschatz zum Teil sehr stark⁶ übereinstimmen (man denke z.B. an Slowenisch – Kroatisch, Tschechisch – Slowakisch usw.). Daher ist im gegebenen Kontext eine weitere Spezifizierung von Internationalismus, nämlich die Berücksichtigung eines griechisch-lateinischen Ursprungs (vgl. dazu Kolwa, 2003, 18), ein aus unserer Sicht sinnvolles spezifizierendes Kriterium.⁷

⁶ Eine ähnliche hohe Gemeinsamkeit könnte sich für romanische und germanische Sprachen ergeben, wenngleich der Grad an Übereinstimmung empirisch zu bestimmen wäre (vgl. dazu eine vielversprechende Untersuchung von Heeringa et al. (2013) zu orthographischen Distanzen in germanischen, slawischen und romanischen Sprachen).

⁷ Eine Spezifizierung für ein Minimalkriterium für Internationalismus wäre das Vorkommen in mindestens drei Sprachen, die wiederum mindestens zwei unterschiedlichen Sprachgruppen/-familien zuzuordnen sind (siehe dazu Krevs Birk & Meulleman, 2024, 334–336).

In jedem Fall ist davon auszugehen, dass Internationalismen beim Spracherwerb eine besondere Rolle spielen, zumal sie das Erlernen und die Rezeption erleichtern. Dies gilt insbesondere für den L2-Spracherwerb, da Internationalismen zumindest das Textverständnis erleichtern können und auch ein aktiver Einsatz durch Lernende zwar vorstellbar ist, aber ab einem bestimmten Kompetenzniveau die Kenntnis der standardsprachlichen Ausdrücke (z.B. *automobil/auto* vs. *kola*, *sekretar* vs. *tajnik*, *plan* vs. *nacrt*) inkl. vorhandener Bedeutungsnuancen notwendig ist. Gleichzeitig können Internationalismen aber auch trotz vorhandener Übereinstimmungen in der Form bei Abweichungen in der Bedeutung zu einer Quelle von Missverständnissen werden (wie z.B. ausführlich u.a. Krevs Birk 2017, 453–456 anhand von Internationalismen im Slowenischen gezeigt werden konnte) und somit als sogenannte „falsche Freunde“ auftreten. Bevor jedoch solche Aspekte untersucht werden können, sollen im nächsten Kapitel unsere empirischen Untersuchungen zu Internationalismen im Kroatischen näher vorgestellt werden. Wie nun zu zeigen sein wird, können bei der Inventarisierung sowohl „objektive“ als auch „subjektive“ Häufigkeiten eine Rolle spielen.

2. „Objektive“ Frequenzen im kroatischen Häufigkeitswörterbuch

Eine kaum lösbare Problematik bei der Ermittlung „objektiver“ Worthäufigkeiten betrifft die Frage des jeweiligen Inputs (= Textbasis) und in der Folge die so genannte Repräsentativität. Deduktiv ist davon auszugehen, dass (1) die Art der ausgewählten Wörter von der zugrunde liegenden Stichprobe abhängt und (2) dies in weiterer Folge auch den Anteil und die Art der Entlehnungen betrifft, die sich auf der Basis eines frequenzbasierten Grundwortschatzes ergeben. Da es sich bei der Lexik um ein offenes und dynamisches System handelt, sind alle Antworten auf (1) und (2) als direkte Konsequenzen der jeweils vorgenommenen Operationalisierung zu sehen. Es stehen prinzipiell viele Möglichkeiten zur Verfügung (Analyse eines bestimmten Autors, eines bestimmten Romans, einer bestimmten Textsorte usw. usf.), und jede empirische Ermittlung von Worthäufigkeiten ist als ein entsprechender „Schnitt“ zu verstehen, dessen Angemessenheit vom jeweiligen linguistischen Erkenntnisinteresse abhängt.

Im Folgenden wird das kroatische *Hrvatski čestotni rječnik* (Moguš et al., 1999, abgekürzt mit *Kro-HWB*) herangezogen. Die Fokussierung auf *Kro-HWB* ist nicht umstritten,⁸ aber als Referenzanalyse aus unserer Sicht durchaus geeignet, um einen strukturierten Einblick in die Frequenzstruktur der kroatischen Lexik zu erhalten. Der Vorteil von *Kro-HWB* liegt nicht im Umfang (der mit einer Million ausgewerteter Wortformen nach heutigem Stand als relativ gering einzuschätzen ist), sondern aus unserer Sicht vor allem in der ausgewogenen Auswahl der jeweiligen „Subkorpora“, die für die Erstellung herangezogen wurden.

⁸ Ein anonymes Reviewer dieses Beitrags weist zu Recht darauf hin, dass für die hier vorgenommenen Analysen auf moderne Webkorpora zurückgegriffen werden sollte. Dieser Hinweis ist durchaus berechtigt, aber – wie im Folgenden gezeigt werden soll – sollte man dem *Kro-HWB* eine berechtigte „Chance“ geben, ausgewertet zu werden. Dies liegt nicht nur daran, dass es für die südslawischen Sprachen generell relativ wenige Häufigkeitswörterbücher gibt (im Vergleich z.B. zum Russischen, vgl. dazu Kelih, 2008, 42), sondern ein großer (und nicht nur heuristischer) Vorteil des *Kro-HWB* besteht darin, dass auch Angaben zum tatsächlichen Vorkommen in einzelnen Funktionalstilen vorliegen. Auch dies wird in diesem Beitrag ausführlich dargestellt. Die Auswertung weiterer Quellen (wie z.B. Webkorpora zu den südslawischen Sprachen) bleibt aber in jedem Fall ein Desiderat.

Es handelt sich jeweils um Texte aus den Bereichen (1) Drama, (2) Journalismus, (3) Prosa, (4) Lyrik und (5) Lehrbücher, wobei sogar eine proportionale Gewichtung der Stichprobengröße vorgenommen wird (vgl. *Kro-HWB*, 6ff. mit Details zu den untersuchten Texten). Nicht von der Hand zu weisen ist allerdings die Problematik, dass hier weniger die traditionellen Funktionalstile zur Verfügung stehen, sondern vor allem Genres aus dem Bereich der literarischen Texte (Drama, Lyrik, Prosa) in das Korpus einfließen und das administrative Register gänzlich ignoriert wird. Der zeitliche Umfang der für das *Kro-HWB* verwendeten Texte beschränkt sich auf das 20. Jahrhundert. Insofern ist chronologisch gesehen aus heutiger Sicht ein älterer Teilbereich der kroatischen Schriftsprache abgedeckt, und durch die selektive Beschränkung auf vor allem literarische, fachsprachliche und journalistische Texte kein Anspruch auf „synchrone“ Repräsentativität erhoben werden kann. In Bezug auf die journalistischen Texte ist bereits an dieser Stelle darauf hinzuweisen, dass diese aus den 70er Jahren des 20. Jahrhunderts stammen, so dass im empirischen Teil der Arbeit gerade diesem Aspekt besondere Aufmerksamkeit zu widmen sein wird.

2.1 Empirische Ergebnisse zu „objektiven“ Frequenzen

Die vorliegende Analyse beschränkt sich, wie bereits erwähnt, auf Internationalismen. Leitendes Kriterium ist ein lateinisch-griechischer etymologischer „Hintergrund“, und es ist davon auszugehen, dass die ermittelten Internationalismen aufgrund ihrer formalen Ähnlichkeit mit dem Deutschen bzw. Englischen kontextunabhängig in ihrer Form „richtig“ dekodiert werden. Es sei explizit darauf hingewiesen, dass dies für die in Frage kommende Bedeutung nicht der Fall sein muss, so dass dieser Aspekt einer gesonderten Untersuchung vorbehalten bleiben muss. Für jedes im *Kro-HWB* vorkommende Wort wird angegeben, in welchen Funktionsstilen (Prosa, Lyrik, Drama, Fachtext, Journalismus) es vorkommt. Dies ermöglicht die Bestimmung der sogenannten Polytextie (vgl. u.a. Köhler, 1986, 92 und insbesondere Baayen, 2010, der allerdings von Dispersion spricht), d.h. des Vorkommens eines Wortes in verschiedenen Textsorten bzw. Funktionalstilen. Damit soll nicht nur die absolute Vorkommenshäufigkeit, sondern eben auch der polyfunktionale Charakter eines Wortes berücksichtigt werden.

Um eine zu kleinteilige Analyse zu vermeiden, werden die im *Kro-HWB* identifizierten Internationalismen im Folgenden zwei Gruppen zugeordnet. Die erste Gruppe sind die so genannten Kern-Internationalismen (im Folgenden als *cor-I* bezeichnet), die in allen oder mindestens vier der untersuchten Funktionsstile vorkommen. Es handelt sich also um Internationalismen, die sowohl eine hohe Vorkommenshäufigkeit als auch eine hohe Polytextie aufweisen. Die zweite Gruppe sind Internationalismen, die als Kriterium in „nur“ drei Funktionalstilen vorkommen und die plakativ als periphere Internationalismen (abgekürzt *perph-I*) bezeichnet werden sollen. Damit ist allerdings keine qualitative Peripherisierung gemeint, sondern eben der eingeschränkte Grad hinsichtlich der Polytextie. Es ist zu erwarten, dass diese innerhalb der hochfrequenten Lexik (= vorderer Rangbereich des *Kro-HWB*, d.h. die Rangplätze 1-1500) eigentlich nur eingeschränkt vorkommen.

Unsere Analyse beschränkt sich auf die 1500⁹ häufigsten Lemmata im Kro-HWB. Innerhalb dieser konnte folgender¹⁰ *cor-I* eruiert werden:¹¹ *organizacija, ekonomski, plan, sistem, problem, funkcija, materijal, turistički, politika, masa, politički, tekst, motor, energija, element, metoda, papir, proces, temperatura, grupa, električan, tehnički, hotel, objekt, automobil, doktor, milijun, nacionalan, organ, aktivnost, socijalistički, program, situacija, akcija, kapetan, revolucija, delegacija, dr., kultura, kemijski, materijalan, film, društveno-politički, cigareta, statistički, vagon, linija, banka, industrija, metal, faktor, karta, kulturan, praksa, tip, dokument, republika, ideja, sekretar, analiza, ton, klasa, kredit, faza, general, minuta, armija, poezija, ministar, tehnika, realan, dolar, scena, baza, karakter, socijalen, turist, organizirati, kriza, tema, fizički, muzika, bomba, kristal, komunist, moderan, komisija und partija*.

Insgesamt lassen sich unter den 1500 häufigsten Wörtern des *Kro-HWB* 88 Internationalismen identifizieren, die sich sowohl durch (1) eine sehr hohe Frequenz als auch durch (2) das Vorkommen in mindestens vier Funktionsstilen (Textsorten) auszeichnen. Das sind immerhin ca. 6% der gesamten untersuchten hochfrequenten Lexik. Davon kommen 47 in allen Funktionalstilen und 41 in vier Funktionalstilen vor, was insgesamt einem hohen Grad an Polytextie gleichkommt. Besonders auffällig ist die absolute Dominanz der Substantive bzw. Adjektive, während im gesamten *cor-I* nur ein einziges Verb (*organizirati*) vorkommt. Diese spezifische Verteilung, d.h. die recht eindeutige Dominanz der Nomen innerhalb der Internationalismen, kann als Hinweis auf eher oberflächliche Spuren lexikalischen Einflusses auf das Kroatische gewertet werden.

Ein näherer Blick auf die eruierten Internationalismen zeigt, dass in der angeführten Liste „markierte“ Items zu finden sind, die zumindest zum Teil, durch die zeitliche Zusammensetzung des verwendeten Korpus, begründet sind, wie z.B. im Falle von *socijalistički, revolucija, delegacija, društveno-politički, sekretar, klasa, armija, komunist, komisija und partija*. Interessanterweise sind diese von uns als „markiert“ angesehenen Wörter aber nicht nur auf die Zeitungssprache beschränkt, sondern sind polyfunktional, zumal sie in zumindest vier Funktionalstilen vorkommen. Für einen Großteil der eruierten Einheiten im *cor-I* ist aber deren Funktionalität im Fremd-

⁹ Die Beschränkung auf die ersten 1500 Lemmata ist wie folgt zu begründen. Erstens soll die Vergleichbarkeit der Ergebnisse mit dem im WOLD-Projekt verwendeten Grundwortschatz (Haspelmath & Tadmor, 2009; Kelih, 2023 für Slowenisch) gewährleistet werden, der ebenfalls ca. 1500 Einträge umfasst. Zweitens entspricht dieser Umfang einem absoluten Minimalwortschatz, wie er in der Fremdsprachendidaktik und Lexikologie diskutiert wird (vgl. Bubenhofer, et al., 2015; Borin, Comrie & Saxena, 2023, 291). Zweifellos hätte aber eine Erweiterung der Stichprobe Auswirkungen auf die Gesamtzahl der Internationalismen.

¹⁰ Insgesamt konnten 115 Internationalismen ermittelt werden, das sind ca. 8% der 1500 häufigsten Lemmata im *Kro-HWB*. Ob dieser Wert hoch oder niedrig ist, lässt sich mangels vergleichbarer Untersuchungen für andere slawische Sprachen nur schwer beurteilen. In Kelih (2025) konnte ein Anteil von 15% Entlehnungen (also nicht nur Internationalismen) im WOLD-Basiswortschatz mit 1500 Bedeutungen festgestellt werden. Die genaue Anzahl der Entlehnungen im Kro-WB müsste allerdings noch ermittelt werden. In Kelih & Vučajnk (2020) konnte in einem Lernerwörterbuch Slowenisch-Deutsch auf der Basis von 4000 Lemmata (Grund- und Aufbauwortschatz) ein Entlehnungsanteil von ca. 23% ermittelt werden. Dieser Wert liegt deutlich über dem Anteil, der in der vorliegenden Analyse für das Kroatische ermittelt werden konnte. Allerdings wäre ein methodologisch korrekter Vergleich nur im Falle möglich, wenn für das Slowenische ein ähnlich zusammengesetztes Häufigkeitswörterbuch wie für das Kroatische zur Verfügung stehen würde bzw. für das Kroatische ebenfalls Lehrbücher bzw. Lernerwörterbücher analysiert würden. In jedem Fall zeigt sich ein erheblicher Forschungsbedarf auf diesem Gebiet. Solche Auswertungen sind sicherlich auch für Erforschung des Purismus im Slowenischen und Kroatischen interessant (vgl. dazu Kalin Golob, 2009 zum Slowenischen, Kunzmann-Müller, 2021, 130–139 überblicksartig zum Kroatischen und insbesondere Stojić & Turk, 2017, 277 ausführlich zum Einfluss von Purismus auf den Umgang mit deutschen Entlehnungen im Kroatischen).

¹¹ Die Anordnung der Lemmata folgt dem Rangplatz, wie im *Kro-HWB* angegeben.

sprachenerwerb auch aus heutiger Perspektive gegeben. In vielen Fällen stellen die Internationalismen ein Reservoir für Synonyme zur Verfügung, wie z.B. im Falle von *glazba/muzika*, *tajnik/sekretar*, *vojska/armija*, *stranka/partija*, *suvremen/moderan*, *metal/kovina* usw., wenngleich hier die vorhandenen Bedeutungsnuancen (wie z.B. bei *vojska/armija* oder *stranka/partija*) gesondert untersucht werden müssten.

Neben dem *cor-I*, der einen polyvalenten Wortschatz darstellt, lassen sich im Hochfrequenz-Bereich des *Kro-HWB* noch einige weitere Internationalismen eruieren, deren Polytextie allerdings eingeschränkt ist. Alle Wörter, die in ≤ 3 Funktionalstilen vorkommen, bilden den *perph-I* mit folgenden Einheiten: *rezultat*, *tom*, *agencija*, *turizam*, *interes*, *magnetski*, *struktura*, *konferencija*, *komitet*, *organizam*, *furnir*, *transformator*, *elektron*, *federacija*, *direktor*, *element*, *delegat*, *reakcija*, *kapacitet*, *kvaliteta*, *sindikata*, *karakteristika*, *spektar*, *biološki*, *opera*, *plastičan* und *industrijski*. Es sind insgesamt 27 Wörter, deren Polytextie eingeschränkt ist, eruierbar. Es ist dies eine relativ heterogene Mischung an Internationalismen, die (zum Teil) kaum der Alltagssprache zuzuordnen sind (wie z.B. *magnetski*, *furnir*, *elektron*, *transformator*, *spektar*), sondern eigentlich als fachsprachlich zu qualifizieren sind. Im Falle von *komitet*, *federacija* und *delegat* ist wiederum die bereits im *cor-I* bemerkte „zeitliche Markierung“ hervorzuheben, zumal zumindest zum Teil direkte Bezüge zur sozialistisch-jugoslawischen politischen Lebenswelt des 20. Jahrhunderts herzustellen sind (z.B. können *federacija*, *komitet* usw. durchaus in aktuellen Kontexten neutral verwendet werden), aber auch dies wäre im Detail erst mit der Hilfe von Assoziations-tests zu überprüfen. In jedem Fall aber erweist sich die Polytextie als eine zusätzliche Möglichkeit die Konsistenz eines Wortschatzes „verlässlicher“ einschätzen zu können, als wenn man nur die objektive Frequenz allein betrachten würde.

Die „objektive“ Häufigkeit ist ein mögliches, aber nicht das einzige Auswahlkriterium für einen Basis- oder Grundwortschatz. Wie in Kapitel 2.2 gezeigt wird, ist aus unserer Sicht auch die „subjektive“ Häufigkeit ein weiteres wertvolles Kriterium. Insgesamt ist festzuhalten, dass Internationalismen im *Kro-HWB* aber durchaus „prominent“ vertreten sind. Zählt man das Inventar von *cor-I* und *perph-I* zusammen, so ergibt sich die Summe von 115 Internationalismen. Bezogen auf das untersuchte hochfrequente Lexikon der 1500 häufigsten Wörter entspricht dies einem Anteil von 7,66%. Ein Großteil der Internationalismen ist in allen Funktionsstilen vertreten und nur ein kleiner Teil (27 Wörter) wird nur in bestimmten Kontexten verwendet. Im Folgenden soll nun der Frage nachgegangen werden, welche Rolle subjektive Sprecherurteile in Ergänzung zur Polytextie und objektiver Frequenz spielen und welche (und eventuell noch feinere) Differenzierung der eruierten Internationalismen möglich ist.

2.2 „Subjektive“ Frequenzen – Ergebnisse zur (synchronen) Geläufigkeit

Aus der psycholinguistischen Literatur ist bekannt, dass es zwischen der sogenannten „objektiven“ Wortfrequenz, die auf der Basis von Texten gewonnen wird, und der sogenannten „subjektiven“ Wortfrequenz auf Basis von Sprecher*innenbefragungen eine hohe Überlappung bzw. Korrelationen zu geben scheint. L1-Sprecher*innen haben eine recht gute Einschätzung über den Frequenzstatus von Wörtern, d.h. ein „Gefühl“ dafür, was sehr oft bzw. was eher selten verwendet wird bzw. geläufig ist.

Für das Kroatische liegt mit „The Croatian psycholinguistic database [...]“ (im Folgenden *CPD*) eine umfangreiche Datenbasis einer sprecher*innenbasierten Erhebung zu Worthäufigkeiten und damit verbundenen Parametern zur Verfügung (vgl. dazu Peti-Stantić 2021 et al.). Für kroatische Substantive, Verben, Adjektive und Adverbien liegen u.a. Daten zu Konkretheit, Bildhaftigkeit, Erwerbsalter und der im gegebenen Kontext relevanten, subjektiven Häufigkeit vor.

Im gegebenen Zusammenhang ist die subjektive Frequenz von Interesse, die auf einer fünfstufigen Skala abgefragt wurde. In der *CPD* liegen für die abgefragten Wörter die jeweiligen Mittelwerte vor, die sich aus der Befragung von über 3500 Probanden ergeben haben. Die gestellte Aufgabe dabei lautete: „[...] procijenite riječi koje slijede na ljestvici od 1 do 5. S riječi za koju me se pita susrećem se:

- 1 = gotovo nikada,
- 2 = jednom godišnje,
- 3 = jednom mjesečno,
- 4 = jednom tjedno,
- 5 = jednom ili više puta dnevno.“

D.h. mit Hilfe dieser Abfrage wird eine bestimmte Art der Bekanntheit bzw. der Geläufigkeit von Ausdrücken auf einer Zeitskala abgefragt. Vom Setting her ist zu erwarten, dass es sich bei einem Grund- bzw. Basiswortschatz um lexikalische Einheiten handeln sollte, mit denen L1-Sprecher*innen ihrer Einschätzung nach täglich oder zumindest wöchentlich konfrontiert werden. Klar ist aber auch, dass die Grenze zu *jednom mjesečno* für die Sprecher*innen schwer zu ziehen ist. In jedem Fall ist aber die Antwort *nie* oder *einmal jährlich* ein starkes Kriterium für den Ausschluss eines Wortes aus einem Grund- bzw. Basiswortschatz.

Im Folgenden soll untersucht werden, welche subjektive Bewertung die von uns in Kapitel 2.1 ermittelten Internationalismen erfahren. Es stellt sich die Frage, ob das auf der Basis „objektiver“ Frequenzen ermittelte Inventar von Internationalismen und ihrer potentiellen Bedeutung für die Kommunikation durch sprecherzentrierte Urteile ihrer Geläufigkeit gestützt werden kann.

Ein erster Anhaltspunkt für die Beantwortung dieser Frage ergibt sich daraus, dass für alle 115 Internationalismen auch die entsprechenden subjektiven Urteile zur Geläufigkeit in der *CPD* vorliegen. Auf der Basis der vorhandenen Mittelwerte¹² der subjektiven Einschätzungen ergibt sich für alle 115 untersuchten Internationalismen ein Mittelwert von 3,61, während sich für das *cor-I*, d.h. diejenigen Internationalismen die sich sowohl durch hohe Vorkommenshäufigkeit als auch hohe Polytextie auszeichnen, ein Mittelwert von 3,75 ergibt. Dies entspricht beides somit tendenziell einer Geläufigkeit auf der obigen Skala, die zwischen einmal pro Monat („jednom mjesečno“) und einmal pro Woche („jednom tjedno“) anzusiedeln ist. Gleichzeitig wird klar, dass hohe Vorkommenshäufigkeit und hohe Polytextie mit einer höheren subjektiven Geläufigkeit einhergehen. Besonders deutlich zeigt sich dies für *perph-I* (d.h. die Internationalismen mit Vorkommen in ≤ 3 Funktionalstilen), dessen mittlere subjektive Geläufigkeit 3,19 beträgt. Damit

¹² De-facto wird hier ein Mittelwert der Mittelwerte aus der *CPD* berechnet. Das ist statistisch erlaubt, aber würde die weitere Anwendung von statistischen Tests auf signifikante Unterschiede zwischen den Gruppen erschweren. Im gegebenen Fall geht es um vor allem auf das Aufzeigen von quantitativen Tendenzen.

zeigt sich, dass die angegebenen Mittelwerte recht gute Indikatoren für die subjektive Gebrauchshäufigkeit sind und *perph-I* tatsächlich offenbar einen begrenzten „Aktionsradius“ hat und dies auch von den Sprecher*innen subjektiv bestätigt wird.

Die weitere Frage, die sich an dieser Stelle stellt, ist, inwiefern sich diese Werte in eine Relation zu den von uns 1500 häufigsten Lemmata aus dem *Kro-HWB* stellen lassen. Aufgrund der Tatsache, dass die eruierten Internationalismen ausschließlich aus Adjektiven und Substantiven bestehen (das einzige Verb *organizirati* soll an dieser Stelle nicht weiter berücksichtigt werden), soll als Vergleichsbasis alle Adjektive und Substantive aus dem Gesamtwortschatz (= 737 Einheiten) herangezogen werden. Berechnet man nämlich für diese spezifische Subgruppe die subjektive Geläufigkeit, so ergibt sich ein Wert von 3,94. Dieser Wert geht eindeutig in Richtung einer relativ hohen Geläufigkeit („jednom tjedno“) und unterscheidet sich von den Ergebnissen zu den Internationalismen, die tendenziell, wie oben gezeigt werden konnte, eine etwas niedrigere Geläufigkeit aufweisen als Nicht-Internationalismen. Dies könnte sich eventuell durch ein geringeres semantisches Potenzial von Entlehnungen erklären, was allerdings erst in gesonderten Studien zu überprüfen wäre. Festzuhalten bleibt aber auch, dass sowohl Internationalismen als auch Nicht-Internationalismen doch recht deutlich von einer Bewertung von 5 (= jednom ili više puta dnevno) entfernt sind. Dies würde man an sich von einem frequenzbasierten Wortschatz durchaus erwarten, was aber ebenfalls in Zukunft noch detaillierter zu untersuchen wäre.

Abschließend gilt es, die sogenannte Polytextie hervorzuheben, die sich als ein durchaus zuverlässiges Kriterium erweist, um die subjektive Geläufigkeit eines Wortes zusätzlich „gewichten“ zu können. Es ist davon auszugehen, dass die subjektive Geläufigkeit eines Wortes u.a. daraus resultiert, dass es in möglichst unterschiedlichen Kontexten vorkommt. Eine mögliche Operationalisierung von Kontext ist in diesem Zusammenhang das Vorkommen von Wörtern in unterschiedlichen Funktionsstilen, wie es im *Kro-HWB* gesondert ausgewiesen wird, was die folgende Analyse erst möglich macht. Im Folgenden soll nun die subjektive Frequenz der Internationalismen nach der Anzahl ihrer Vorkommen in Funktionalstilen (in einem, zwei, drei ...) näher betrachtet werden. Wie die Abbildung 1 zeigt, ergibt sich aus dieser Perspektive ein recht eindeutiger statistischer Zusammenhang zwischen Polytextie und subjektiver Häufigkeit, der besagt: Je höher die Anzahl der Funktionalstile, in denen ein Internationalismus vorkommt, desto höher ist auch die subjektive Einschätzung durch die Sprecher*innen.

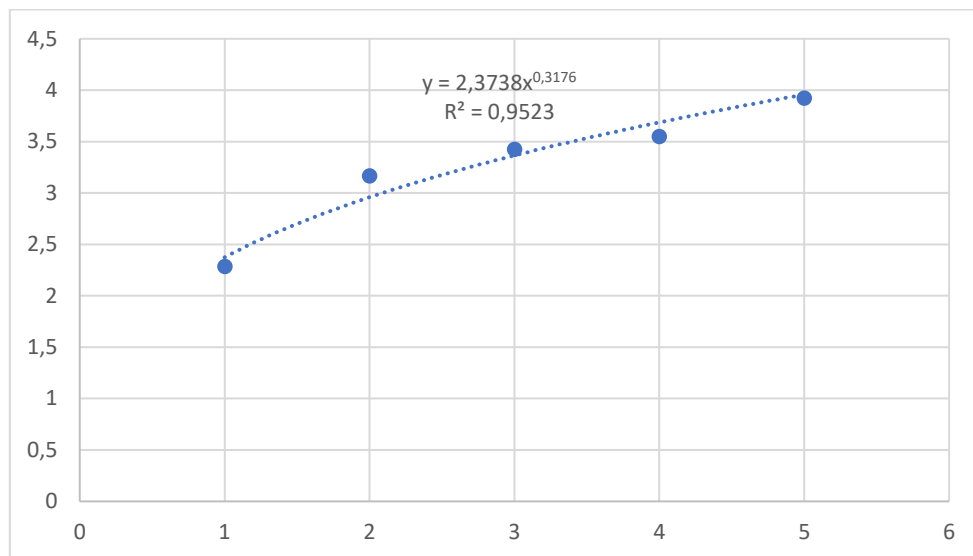


Abb. 1: Polytextie (x-Achse) vs. durchschnittliche „subjektive“ Frequenz (y-Achse)

Dieser Zusammenhang lässt sich auch mit einem einfachen nichtlinearen Potenzmodell erfolgreich modellieren (mit $R^2 = 0,95$; das Modell ist in Abb. 1 angeführt), was insgesamt auf eine starke Wechselwirkung zwischen den genannten Variablen hindeutet. Damit wird deutlich, dass sowohl die Polytextie als auch die subjektive Häufigkeit nicht nur wichtige komplementäre Eigenschaften zur objektiven Häufigkeit sind, sondern dass auf der Basis des obigen Modells die subjektive Einschätzung ausgehend von der Polytextie in gewisser Weise sogar vorhersagbar ist. Der praktische Nutzen aus dem obigen Modell lässt sich abschließend gut daran illustrieren, dass die von uns im *cor-I* identifizierten und dort als „markiert“ bezeichneten Wörter (*socijalistički, revolucija, delegacija, društveno-politički, sekretar, klasa, armija, komunist, komisija, partija*) tatsächlich über einen recht niedrigen Mittelwert von 2,90 durchschnittlicher Geläufigkeit verfügen, was laut dem obigen Modell eindeutig auf einen eingeschränkten Verwendungskontext hinweist. Nun kann das in Kap. 2.1 eruierte *perph-I* (das Kriterium mit dem Vorkommen in ≤ 3 Funktionalstilen ergab dieses Inventar: *rezultat, tom, agencija, turizam, interes, magnetski, struktura, konferencija, komitet, organizam, furnir, transformator, elektron, federacija, direktor, elemenat, delegat, reakcija, kapacitet, kvaliteta, sindikat, karakteristika, spektar, biološki, opera, plastičan* und *industrijski*) durch die subjektive Einschätzung „gewichtet“ werden. Betrachtet man diejenigen Wörter, die einen Mittelwert $> 3,75$ aufweisen (= *rezultat, interes, struktura, direktor, reakcija, kvaliteta, karakteristika*), sind diese auf dieser Basis als durchaus zum *cor-I* gehörige Einheiten aufzufassen. D.h. es bietet sich eine Kombination von objektiver Frequenz und subjektiver Einschätzung und der Polytextie als sinnvolles, intersubjektiv nachvollziehbares Set von Eigenschaften an, um im Bedarfsfall intersubjektiv nachvollziehbare Selektionskriterien zur Hand zu haben. Es sei jedoch angemerkt, dass vor weiterführenden Verallgemeinerungen eine Ergänzung des vorliegenden Häufigkeitswörterbuchs durch die Verwendung von modernen (Web-)Korpora empfehlenswert ist. Dies dient der Erlangung eines empirisch soliden Bildes sowie der Bestätigung der bislang gewonnenen Resultate.

3. Zusammenfassung

Zweifellos sind sowohl die objektive als auch die subjektive Frequenz Merkmale, die sowohl aus linguistischer als auch aus fremdsprachendidaktischer Perspektive von Bedeutung sind. In beiden Fällen handelt es sich um Konzepte, die je nach Operationalisierung unterschiedliche empirische Realisierungen erfahren. Insofern sind alle Ergebnisse und Interpretationen in diesem Licht zu sehen. Im vorliegenden Fall liegt der Fokus auf der hochfrequenten Lexik des Kroatischen mit einem Inventar von 1500 Wörtern. Das verwendete *Kro-HWB* hat in Hinblick auf den Input (Fokussierung auf bestimmte Genres bzw. Funktionalstile) und den Stichprobenumfang (zwar balanciert, aber quantitativ nicht mit modernen Webkorpora zu vergleichen) durchaus seine konzeptuellen Schwächen, ermöglicht aber eine systematische Untersuchung der sogenannten Polytextie, die, wie gezeigt werden konnte, je nach Bedarf und Forschungsinteresse ergänzend zur objektiven bzw. subjektiven Frequenz heranzuziehen ist. Zu beachten ist auch die zeitliche Zusammensetzung des Textinputs (der Schwerpunkt der in das Korpus eingehenden Texte ist auf dem 20. Jahrhundert), die sich auch im hochfrequenten Bereich der Lexik niederschlägt. Dies lässt sich jedenfalls aus einem Teil des ermittelten Inventars an Internationalismen erkennen, wenngleich gerade bei der Ermittlung einer „markierten“ Lexik nicht die eigene Intuition, sondern die Polytextie und die subjektiven Häufigkeitseinschätzungen herangezogen werden sollten. Diese sind aus unserer Sicht ein sinnvolles Instrumentarium (klarerweise wiederum in Abhängigkeit von der Repräsentativität, Streuung nach Alter, dem sozialen Status, Wohnort usw.), um die Geläufigkeit bzw. die Bekanntheit von Lexemen besser einschätzen zu können. Darüber hinaus konnte gezeigt werden, dass subjektive Sprecherurteile sich in einer modellierbaren Wechselbeziehung zur Polytextie befinden. Auch wenn demnach objektive bzw. subjektive Frequenz und Polytextie als korrelierte Größen zu betrachten sind, spielen sie ihre Stärke, wie in den obigen Analysen gezeigt werden konnte, vor allem in Kombination aus. Abschließend ist auf das auf intersubjektive Weise gewonnene Inventar von Internationalismen zu verweisen, das sowohl aus linguistischer als auch aus fremdsprachendidaktischer Perspektive von Bedeutung ist. Diese sind ein nützliches Instrument (natürlich wieder in Abhängigkeit von Repräsentativität, Streuung nach Alter, sozialem Status, Wohnort etc). Auch wenn also objektive bzw. subjektive Frequenz und Polytextie als korrelierte Größen zu betrachten sind, spielen sie, wie in den obigen Analysen gezeigt werden konnte, ihre Stärke vor allem in Kombination aus. So sei abschließend auf das intersubjektiv gewonnene Inventar der Internationalismen verwiesen.

Zweifellos ist die Kenntnis der hochfrequenten Internationalismen nicht nur eine kognitive „Erleichterung“ im Spracherwerb, sondern ihr Anteil ist an sich von Interesse für die empirische Sprachkontaktforschung. Der ermittelte Anteil von ca. 8% Internationalismen lässt eine gute Positionierung des üblicherweise als puristisch bezeichneten Kroatisch zu. Bedenkt man, dass darüber hinaus auch andere Arten von lexikalischen Entlehnungen unter den hochfrequenten Lexemen zu finden sind, so erscheint es durchaus plausibel, dass etwa ein Sechstel des Gesamtinventars als Entlehnungen im weiteren Sinne anzusehen ist. Dies ist insgesamt keine vernachlässigbare Größe. Generell lässt sich somit sagen, dass die hochfrequente Lexik des Kroatischen keine besonders ausgeprägte Entlehnungsresistenz aufweist.

Daraus ergibt sich perspektivisch, dass es zwischen der frequenzbasierten Wortschatzauswahl und dem in der historisch-vergleichenden Sprachwissenschaft diskutierten Basiswortschatz hinsichtlich der darin enthaltenen Entlehnungen gewisse Überschneidungen geben könnte. In beiden Fällen geht es implizit im weitesten Sinne um einen Wortschatz, der eine hohe Relevanz für die Bewältigung kommunikativer Aufgaben im Alltag haben sollte, wobei aus erkenntnistheoretischen Gründen „empirisch gewichteten“ Ansätzen der Vorzug zu geben ist.

Literaturverzeichnis

- Anstatt, T. (2016). Subjektive Frequenz als Forschungsmethode. *Wiener Slawistischer Almanach* 77, 7–35.
- Backus, A. (2013). A usage-based approach to borrowability. In E. Zenner & G. Kristiansen (Hrsg.), *New Perspectives on Lexical Borrowing. Onomasiological, Methodological and Phraseological Innov* (19–39). De Gruyter Mouton.
- Baayen, H. (2010). Demythologizing the word frequency effect. A discriminative learning perspective. *The Mental Lexicon* 5(3), 436–461. DOI: 10.1075/ml.5.3.10baa
- Baayen, H., Feldman, L. & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 55(2), 290–313. DOI: 10.1016/j.jml.2006.03.008
- Balota, D., Pilotti, M. & Cortese, M. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition* 29(4), 639–647.
- Borin, L., Comrie, B. & Saxena, A. (2013). The Intercontinental Dictionary Series – A rich and principled database for language comparison. In A. Saxena & L. Borin (Hrsg.), *Approaches to measuring linguistic differences* (285–302). De Gruyter Mouton.
- Braun, P. (1990). Internationalismen – Gleiche Wortschätze in europäischen Sprachen. In P. Braun, B. Schaefer & J. Volmert (Hrsg.), *Internationalismen. Studien zur interlingualen Lexikologie und Lexikographie* (13–33). Niemeyer.
- Bubenhofer, N., Lange, W., Okamura, S. & Scharloth, J. (2015). Wortschätze in Lehrbüchern für Deutsch als Fremdsprache – Möglichkeiten und Grenzen frequenzorientierter Ansätze. In J. Kiesendahl & Ch. Ott (Hrsg.), *Linguistik und Schulbuchforschung. Gegenstände – Methoden – Perspektiven* (85–108). V&R Unipress.
- Glück, H. (2010). *Metzler Lexikon Sprache*. 4. aktualisierte und überarbeitete Auflage. Metzler.
- Haderlein, V. (2008). *Das Konzept zentraler Wortschätze. Bestandsaufnahme, theoretisch-methodische Weiterführung und praktische Untersuchung*. [=unveröffentlichte Dissertation]. LMU München.
- Haspelmath, M. & Tadmor, U. (Hrsg.) (2009). *Loanwords in the world's languages. A comparative handbook*. De Gruyter Mouton.
- Heeringa, W., Golubovic, J., Gooskens, Ch., Schüppert, A., Swarte, F. & Voigt, St. (2013). Lexical and orthographic distances between Germanic, Romance and Slavic languages and their relationship to geographic distance. In Ch. Gooskens & R. van Bezooijen (Hrsg.), *Phonetics in Europe. Perception and Production* (99–137). Lang.
- Kalin Golob, Monika (2009). Linguistic purism in Slovene language: From Trubar to the present. In J. Granić (Hrsg.), *Jezična politika i jezična stvarnost* (137–146). Hrvatsko društvo za primijenjenu lingvistiku.
- Kelih, E. (2008). *Geschichte der Anwendung quantitativer Verfahren in der russischen Sprach- und Literaturwissenschaft*. Kovač.
- Kelih, E. (2023). *Entlehnungen im slowenischen Basiswortschatz – ein gebrauchsbasierter Ansatz*. Lang.
- Kelih, E. (2025). Entlehnungen im kroatischen Basiswortschatz (WOLD-Liste) [eingereicht, *Anzeiger für Slavische Philologie*]
- Kelih, E. & Vučajnk, T. (2020). Prevzete besede v slovenskem osnovnem besedišču (prva opažanja). In M. Šekli & L. Rezončnik (Hrsg.), *Slovenski jezik in književnost v srednjeevropskem prostoru* (61–72). Slavistično društvo Slovenije.
- Köhler, R. (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Brockmeyer.
- Köhler, R. (2005a). Properties of lexical units and systems. In R. Köhler, G. Altmann & R.G. Piotrowski (Hrsg.), *Quantitative Linguistics. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook* (760–774). De Gruyter.

- Köhler, R. (2005b). Gegenstand und Arbeitsweise der Quantitativen Linguistik. In R. Köhler, G. Altmann & R. G. Piotrowski (Hrsg.), *Quantitative Linguistik. Quantitative Linguistics* (1–16). De Gruyter.
- Kolwa, A. (2003). Zur Geschichte der Internationalismen-Forschung. In P. Braun, B. Schaefer & J. Volmert (Hrsg.), *Internationalismen II. Studien zur interlingualen Lexikologie und Lexikographie* (13–21). De Gruyter.
- Kreys Birk, U. (2017). Zu einigen aktuellen Reflexen des deutsch-slowenischen Sprachkontaktes. In Ch. Mauerer (Hrsg.), *Mehrsprachigkeit in Mittel-, Ost- und Südosteuropa. Gewachsene Vielfalt oder belastendes Erbe der Vergangenheit: Beiträge zur 1. Jahrestagung des Forschungszentrums Deutsch in Mittel-, Ost- und Südosteuropa, Regensburg, 2.-4. Oktober 2014* (449–469). Pustet.
- Kreys Birk, U., Meullemann, M. (2024). Lehnwörter und Internationalismen aus dem Französischen im Slowenischen: die Vermittlerrolle des Deutschen. *Germanistische Beiträge* 50(1), 321–341. DOI: 10.2478/gb-2024-0018.
- Kunzmann-Müller, B. (2021) *Lexikologie der kroatischen Sprache der Gegenwart. Eine Einführung. Unter Mitarbeit von Ana Nemec*. Buske.
- Levickij, V.V. (2005). Polysemie. In R. Köhler, G. Altmann & R.G. Piotrowski (Hrsg.), *Quantitative Linguistik. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook* (458–464). De Gruyter.
- Krohn, D. (1992). *Grundwortschätze und Auswahlkriterien. Metalexikographische und fremdsprachendidaktische Studien zur Struktur und Funktion deutscher Grundwortschätze*. Acta Univ. Gothoburgensis.
- McGee, I. (2006). *Lexical Intuitions and Collocation Patterns in Corpora*. Cardiff University.
- Moguš, M., Bratanić, M. & Tadić, M. (1999). *Hrvatski čestotni rječnik*. Školska knjiga.
- Peti-Stantić, A., Anđel, M., Gnjidić, V., Keresteš, G., Ljubešić, N., Masnikosa, I., Tonković, M., Tušek, J., Willer-Gold, J. & Stanojević, M.-M. (2021). The Croatian psycholinguistic database: Estimates for 6000 nouns, verbs, adjectives and adverbs. *Behavior research methods* 53(4), 1799–1816.
- Pustet, R. (2004). Zipf and his heirs. *Language Sciences* 26(1), 1–25.
- Schaefer, B. (1990). Versuch einer theoretischen Grundlegung der Internationalismen-Forschung. In P. Braun, B. Schaefer & Volmert, J. (Hrsg.), *Internationalismen. Studien zur interlingualen Lexikologie und Lexikographie* (34–46). Niemeyer.
- Schaefer, B. (2003). Neuerlicher Versuch einer theoretischen und methodischen Grundlegung der Internationalismen-Forschung. In P. Braun, B. Schaefer & J. Volmert (Hrsg.), *Internationalismen II. Studien zur interlingualen Lexikologie und Lexikographie* (71–107). De Gruyter.
- Stefanowitsch, A. (2020). *Corpus linguistics. A guide to the methodology*. Language Science Press.
- Stojić, A., Turk, M. (2017). *Deutsch-kroatische Sprachkontakte. Historische Entwicklung und aktuelle Perspektiven auf lexikalischer Ebene*. Narr Francke Attempto.
- Tadmor, U., Haspelmath, M. & Taylor, B. (2010). Borrowability and the notion of basic vocabulary. *Diachronica* 27(2), 226–246.
- Thomason, S. G. (2001). *Language contact*. Edinburgh University Press.
- Tuldava, J. (1998). *Probleme und Methoden der quantitativ-systemischen Analyse*. Wissenschaftlicher Verlag Trier.
- Zenner, E., Speelman, D. & Geeraerts, D. (2014). Core vocabulary, borrowability and entrenchment: A usage-based onomasiological approach. *Diachronica* 31(1), 74–105.
- Ziem, A. & Lasch, A. (2013). *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze*. De Gruyter.