

Inhalt

Cyril Robert Brosch & Sabine Fiedler <i>Einführung</i>	7
Věra Barandovská-Frank <i>Mögliche Szenarien for the future of Interlinguistics</i>	9
Pierre Dieumegard & Chloé Izquierdo <i>Zur Bekanntheit konstruierter Sprachen in Frankreich: Ergebnisse einer Umfrage im Jahr 2022</i>	27
Sabine Fiedler <i>Interlinguistik for future? – Einige Überlegungen zu Zielen und zur Zukunft der Interlinguistik</i>	39
Cornelia Mannewitz <i>Ortsnamen in der Ukraine</i>	49
Klaus Schubert <i>Interlinguistik im 19. Jahrhundert: Die Société internationale de Linguistique und die Universalsprachen</i>	55
Christian Siefkes <i>Lugamun – eine mit algorithmischer Unterstützung erstellte »Worldlang«</i>	105
Sabine Fiedler <i>Buchvorstellung: Erich-Dieter Krause: Vollständiges Wörterbuch Deutsch-Esperanto (in drei Bänden)</i>	127
Goro Christoph Kimura <i>Buchvorstellung: Ermunterung zum sparsameren Gebrauch des Englischen</i>	137
<i>Über die Autoren</i>	145
Akten der Gesellschaft für Interlinguistik	147

Cyril Robert Brosch & Sabine Fiedler

Einführung

Das *Jahrbuch der Gesellschaft für Interlinguistik* des Jahres 2023 bietet einen Überblick über Vorträge, die auf der letzten Jahrestagung gehalten wurden, welche vom 18. bis 20. November 2022 in Berlin mit dem Schwerpunktthema »Interlinguistik *for future*« stattfand. Wie für das GIL-Jahrbuch üblich, enthält es darüber hinaus aber auch weitere Aufsätze, die im zu Ende gehenden Jahr entstanden und unter dem genannten Thema relevant sind. Die Beiträge seien in der Folge kurz vorgestellt:

Věra Barandovská-Frank eröffnet den Band mit dem Beitrag »Mögliche Szenarien for the future of Interlinguistics«, in dem sie zunächst zeigt, wie das Feld der Interlinguistik seit Einführung des Ausdrucks 1911 beständig erweitert wurde (von der Kunst des Entwurfs von Plansprachen hin zu einer Disziplin, die alle Facetten internationaler Sprachkontakte abdeckt), um dann vier mögliche künftige Entwicklungslinien zu nennen, gemäß denen sich das Fach künftig entwickeln könnte.

Pierre Dieumegard und **Chloé Izquierdo** stellen in »Zur Bekanntheit konstruierter Sprachen in Frankreich: Ergebnisse einer Umfrage im Jahr 2022« anhand zahlreicher Schaubilder dar, welche Plansprachen (sowohl internationale Plansprachen als auch Kunstsprachen aus Film und Literatur) einer relativ großen Stichprobe von Personen (in welchem Umfang) bekannt sind. In den zwei Gruppen stechen Esperanto einerseits und Tolkiens Elbensprachen andererseits deutlich hervor.

Sabine Fiedler reflektiert in ihrem Beitrag »Interlinguistik *for future?* – Einige Überlegungen zu Zielen und zur Zukunft der Interlinguistik«, ähnlich wie Barandovská-Frank, darüber, ob die Interlinguistik künftig einen noch stärkeren Fokus auf Kunstsprachen wie Klingonisch oder Dothraki haben wird. Sie plädiert für vermehrte Aufklärung über deren Unterschiede zu internationalen Plansprachen wie Esperanto und eine stärkere interdisziplinäre Zusammenarbeit zwischen der Interlinguistik und anderen Fächern.

Von besonderer, keiner Erklärung bedürftiger Aktualität ist der Beitrag von **Cornelia Mannewitz** »Ortsnamen in der Ukraine«, der die heterogene Herkunft und verschlungene Geschichte der ukrainischen Toponyme aufzeigt. Gerade das Spannungsfeld zwischen russischen und ukrainischen Elementen, oft im selben Wort, zeigt, dass auch dieses Thema zur weit gefassten Interlinguistik (als Wissenschaft vom Sprachkontakt) zu rechnen ist.

Der besonders umfangreiche Beitrag von **Klaus Schubert** »Interlinguistik im 19. Jahrhundert: Die Société internationale de Linguistique und die Universalsprachen« dokumentiert anhand der Originalquellen die Arbeit einer Pariser Sprachgesellschaft, die, anders als ihre bekannteren, aber organisatorisch und personell unabhängigen Namensvettern, um die Mitte des 19. Jahrhunderts apriorische Plansprachenprojekte diskutierte und bewertete, wie man ihrer kurzlebigen Zeitschrift entnehmen kann.

Schließlich stellt **Christian Siefkes** »Lugamun – eine mit algorithmischer Unterstützung erstellte ›Worldlang‹« vor, ein System einer internationalen Plansprache neueren Typs, deren

(lexikalische) Quellen nicht auf einige ursprünglich (west-)europäische Sprachen beschränkt sind. Die besondere Herausforderung bei einer solchen heterogenen Grundlage ist allerdings die Wiedererkennbarkeit des Wortschatzes, die durch einen Computeralgorithmus (mit menschlicher Nachkontrolle) erreicht wird.

Den Abschluss des Jahrbuches 2023 bilden Informationen zu zwei interlinguistisch relevanten Publikationen. **Sabine Fiedler** stellt Erich-Dieter Krauses »Vollständiges Wörterbuch Deutsch-Esperanto (in drei Bänden)«, das umfangreichste zweisprachige Wörterbuch des Esperanto aller Zeiten, vor. Das Wörterbuch wird im Detail präsentiert und vor dem Hintergrund moderner lexikologischer Anforderungen einer kritischen Durchsicht unterzogen. Daran schließt sich die Buchvorstellung von **Goro Christoph Kimura** an, der sein eigenes, in Japanisch verfasstes Werk »Ermunterung zum sparsameren Gebrauch des Englischen« dem deutschsprachigen Leserkreis zugänglich macht. Ganz dem Titel entsprechend wird empfohlen, die heute teils dominierende Verwendung des Englischen als Verkehrssprache zugunsten anderer, im Buch vorgestellter Strategien zu reduzieren, gerade um die Herausforderungen der Internationalisierung und Globalisierung besser zu meistern.

Mit der hier präsentierten Mischung aus fachtheoretisch ausgerichteten Artikeln zur Weiterentwicklung der Interlinguistik, empirisch basierten Studien zu spezifischen interlinguistischen Fragestellungen und der Vorstellung neuer Publikationen aus dem Fachbereich vermittelt das Jahrbuch der GIL in diesem Jahr in ganz besonderem Maße einen Einblick in die große Breite aktueller interlinguistischer Forschung.

Alle Bände des JGI sowie seinem Vorgängerformat »Beihefte zu den Interlinguistischen Informationen« (s. hier ab Seite 147) können kostenlos unter www.interlinguistik-gil.de heruntergeladen werden.

Christian Siefkes

Lugamun – eine mit algorithmischer Unterstützung erstellte »Worldlang«

A comparatively new and small category of international auxiliary languages are the so-called "worldlangs", whose vocabulary and grammar are based not only on the Indo-European languages but also on languages from other language families. But the more diverse the source languages are, the more difficult it is to justify the decisions on the use of certain words, grammatical structures, etc. Why use one particular variant, when another might have been as good or better? Lugamun is a new worldlang that uses an algorithm to ease this problem, based on free online sources such as Wiktionary, the multilingual dictionary. For each term to be added to its dictionary, the words used in the various source languages are automatically converted to fit Lugamun's spelling and phonology, and then automatically ranked according to several criteria, such as their similarity to the words used in other languages. By default, the best-ranked candidate is then added to Lugamun's vocabulary; another candidate can be chosen if preferred, but such choices must always be justified. In this way it is possible to understand the reasoning behind the choice of each word. A similarly systematic process is also used to select of grammatical structures.

Relative nova kaj malgranda kategorio de internaciaj helplingvoj estas la "mondfontaj planlingvoj", kies vortprovizo kaj gramatiko baziĝas ne nur sur la hindeŭropaj lingvoj, sed ankaŭ sur lingvoj el aliaj familioj. Sed ju pli diversaj la fontlingvoj, des pli malfacilas pravigi la decidojn uzi certajn vortojn, gramatikajn strukturojn ktp. Kial uzi certan varianton, se alia povus esti same bona aŭ pli bona? Lugamun estas nova mondfonta planlingvo, kiu uzas algoritmon por faciligi ĉi tiun problemon surbaze de liberaj interretaj informaroj, kiel la multalingva Viki-vortaro. Por ĉiu vorto aldonota al la vortaro, la vortoj uzataj en la diversaj fontlingvoj estas aŭtomate konvertitaj al la ortografio kaj fonologio de Lugamun, kaj aŭtomate vicigitaj surbaze de pluraj kriterioj, kiel ilia simileco al la vortoj uzataj en aliaj lingvoj. Ĝenerale la plej bone vicigita kandidato estas aldonita al la lugamuna vortaro; alia kandidato povas esti elektita se preferinda, sed tiaj elektoj estas ĉiam skribe pravigitaj. Tiel eblas kompreni por ĉiu vorto la motivon je ĝia elekto. Simile sistema procezo ankaŭ estas uzata por la elekto de gramatikaj strukturoj.

1 Lugamun, eine »Worldlang«

Lugamun ist eine Abkürzung von *luga komun*, was auf Lugamun »gemeinsame Sprache« bedeutet. Schon anhand der Auflösung seines Namens lässt sich erkennen, dass Lugamuns Wortschatz verschiedenen Sprachfamilien entstammt: *luga* kommt aus dem Arabischen und Swahili, *komun* hingegen aus den romanischen Sprachen und dem Englischen. Tatsächlich ist es eine als internationale Hilfssprache konzipierte Plansprache der Art, die als »Worldlang« bezeichnet wird. Mit diesem Begriff (Esperanto: *mondfonta planlingvo*) werden Plansprachen bezeichnet, deren Vokabular und Grammatik nicht bloß auf indoeuropäischen (oder gar bloß westeuropäischen) Sprachen aufbauen, sondern auch Sprachen aus anderen Sprachfamilien einbeziehen (IAL Wiki 2019).

Plansprachen wie Esperanto, die sich in Grammatik und Vokabular überwiegend auf europäische Sprachen stützen, haben den Vorteil, dass diese Sprachen viele Gemeinsamkeiten haben, auf die die Plansprache zurückgreifen kann. Doch je vielfältiger und je weniger verwandt die Quellsprachen, desto mehr stellt sich die Problematik, wie die Entscheidungen für bestimmte

Wörter, grammatische Strukturen usw. begründet werden können – warum gerade diese Variante, wäre nicht eine andere ebenso gut oder besser gewesen?

Im Folgenden soll, insbesondere in Bezug auf Phonologie und Vokabular, dargestellt werden, wie diese Herausforderung in Lugamun angegangen wird. Auf die Grammatik kann nur kurz eingegangen werden, doch einige wesentliche Elemente werden anhand eines Textbeispiels dargestellt. Schließlich wird noch kurz die Geschichte und die (eher unsichere) Zukunft Lugamuns thematisiert.

2 Quellsprachen

Während Esperanto und auch die meisten anderen Plansprachen fast ausschließlich auf europäischen Sprachen basieren, hat Lugamun den Anspruch, die Vielfalt der in der Welt gesprochenen Sprachen besser abzubilden. Dabei ergibt sich allerdings ein unvermeidliches Problem: Aus je mehr Quellsprachen das Vokabular einer Plansprache kommt, desto unvertrauter wird es einem Sprecher irgendeiner dieser Sprachen sein. Als Extremfall könnte man sich eine Sprache vorstellen, die jeder der über 7000 weltweit gesprochenen Sprachen ein Wort entnimmt. Dann würde jeder Mensch ein Wort dieser Sprache aus seiner eigenen kennen – aber eben nur eins. Oder einige wenige, wenn die Person noch ein paar andere Sprachen spricht. Aber der Rest müsste komplett neu gelernt werden, wie es auch bei apriorischen Plansprachen wie Kotava (Fetcey & Comité Linguistique Kotava 2022) der Fall ist. Die Sprache wäre damit zwar sehr »neutral«, aber auch schwieriger zu lernen als eine Sprache, deren Vokabular viele Überlappungen mit den Muttersprachen der Lernenden aufweist.

Eine weitere Hürde bei der Berücksichtigung vieler Quellsprachen ist es, dass der Rechercheaufwand mit der Berücksichtigung weiterer Quellsprachen linear ansteigt, jedenfalls dann, wenn bei der Auswahl jedes neuen Begriffs möglichst die entsprechenden Wörter aus *allen* Quellsprachen berücksichtigt werden sollen – was bei Lugamun der Fall ist. Bei der Erstellung von Lugamun wird daher bewusst nur eine relativ kleine Anzahl an Quellsprachen berücksichtigt, nämlich zehn. Zwar können so längst nicht alle Einzelsprachen berücksichtigt werden, doch sind die Quellsprachen sorgfältig ausgewählt, um zumindest die großen Sprachfamilien abzubilden.

Die Hälfte dieser Sprachen – nämlich Englisch, Hindi/Urdu, Spanisch, Französisch und Russisch – gehört zur indoeuropäischen Sprachfamilie, der Sprachfamilie mit der bei weitem höchsten Zahl von Sprechern. Die anderen fünf Quellsprachen – Mandarin (Hochchinesisch), Arabisch, Indonesisch/Malaiisch, Japanisch und Swahili – repräsentieren fünf weitere besonders weit verbreitete Sprachfamilien. Da indoeuropäische Sprachen so extrem weit verbreitet sind – sie werden von etwa 40 Prozent der Weltbevölkerung gesprochen – wurden hier mehrere Quellsprachen ausgewählt, die jeweils die größten Zweige dieser Familie repräsentieren.

Für jeden Zweig, der mindestens von hundert Millionen Menschen gesprochen wird, wurde dabei die am weitesten verbreitete Sprache als Vertreter gewählt: Englisch für die germanischen, Hindi/Urdu für die indoiranischen, Spanisch für die romanischen und Russisch für die slawischen Sprachen. Französisch, die romanische Sprache mit der zweitgrößten Zahl an Sprechenden, taucht in dieser Liste nicht auf. Es wurde trotzdem als weitere Quellsprache ausgewählt,

da es extrem weit verbreitet ist. Es ist in mehr als 30 Ländern Amtssprache (die zweithöchste Zahl nach dem Englischen). Zudem ist es eine der Amtssprachen der Vereinten Nationen, und seine Berücksichtigung führt dazu, dass *alle* UN-Amtssprachen auch Quellsprachen sind.

Vier der weiteren Quellsprachen stehen jeweils als Repräsentanten – als die Sprache mit der größten Sprecherzahl – für die vier nächstgrößten Sprachfamilien: Hochchinesisch für die sinotibetischen, Arabisch für die afroasiatischen, Indonesisch für die austronesischen und Swahili für die Niger-Kongo-Sprachen. Die letzte Quellsprache – Japanisch – ist die größte Sprache (nach Sprecherzahl) aus einer Familie, die anderweitig noch nicht in den Quellsprachen vertreten ist, nämlich der japanischen Sprachfamilie (auch Japanisch-Ryūkyū genannt).

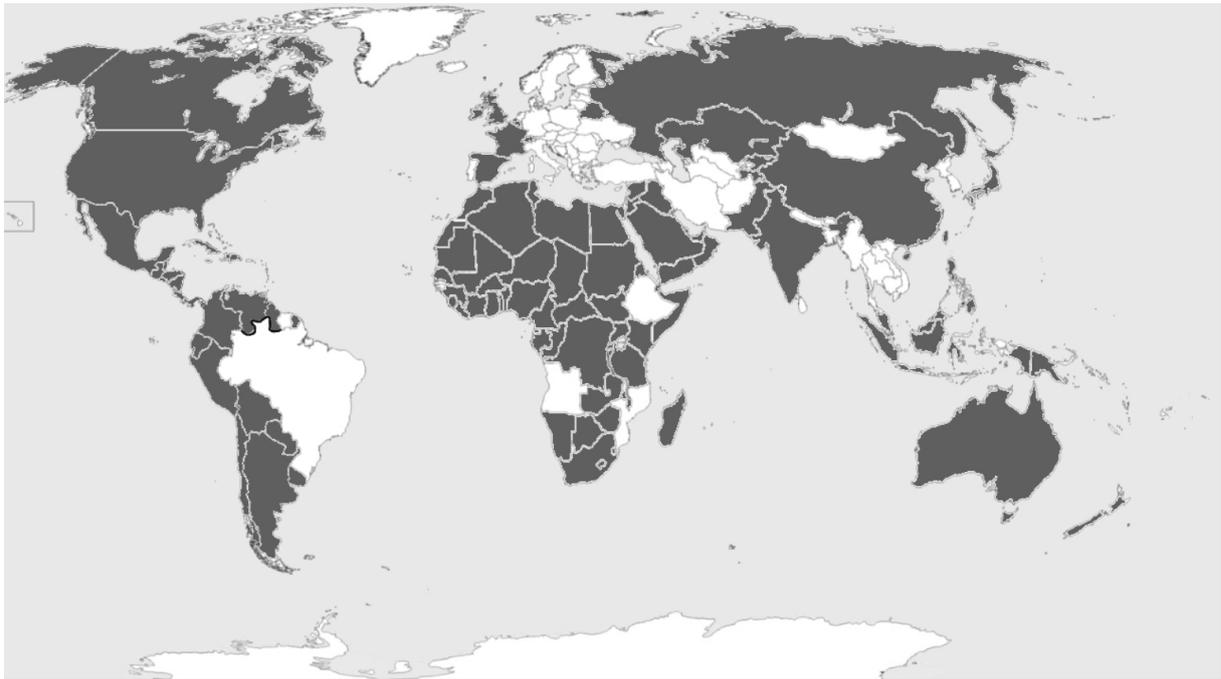


Abbildung 1: Länder, die eine der Quellsprachen als Amtssprache haben

Abb. 1 zeigt die Länder, in denen eine der Quellsprachen Amtssprache ist. Wie man sieht, ist ein Großteil der Welt schon direkt vertreten. Berücksichtigt man zudem, dass jede der Sprachen stellvertretend für eine ganze Familie steht, schließen sich viele der hier noch weiß gefärbten Lücken. So wird etwa in Brasilien, Angola und Mosambik das dem Spanischen eng verwandte Portugiesisch gesprochen; in Iran, Afghanistan und Bangladesch werden indoiranische Sprachen gesprochen; in Myanmar eine sinotibetische. Auch wenn zwangsläufig nicht jede Sprache und Sprachfamilie berücksichtigt werden kann, wird so doch eine beachtliche Abdeckung erreicht.

3 Phonologie

Ein Prinzip im Design von Lugamun ist es, eine möglichst »durchschnittliche« Sprache zu sein, deren Merkmale möglichst vielen Menschen aus ihren eigenen Sprachen schon bekannt sein sollen. Dafür orientieren wir uns insbesondere an wissenschaftlichen Sammlungen von »typischen« grammatischen und phonetischen Merkmalen. Besonders nützlich ist WALS, der *World*

Atlas of Language Structures – eine Onlinequelle, die für fast 200 Merkmale zusammengetragen hat, in welchen Sprachen diese in welcher Ausprägung vorhanden sind (Dryer & Haspelmath 2013). So enthält WALs (Kap. 2) die Information, dass die meisten Sprachen fünf bis sechs Vokale haben – der Medianwert (Zentralwert) wird nicht genannt, aber der Durchschnitt ist weniger als sechs.

	vorne	zentral	hinten
geschlossen	i		u
mittel	e		o
offen	a		

Abbildung 2: Die Vokale von Lugamun

Um nicht »überdurchschnittlich« komplex zu sein, begnügt sich Lugamun daher mit fünf Vokalen, und zwar mit denjenigen, die auch am häufigsten in anderen Sprachen zu finden sind. Dafür haben wir eine andere Onlinequelle konsultiert: PHOIBLE 2.0 (Moran & McCloy 2019). Diese Online-Datenbank sammelt Informationen darüber, welche Phoneme (in IPA-Notation) in welchen Sprachen zu finden sind und ermöglicht dabei auch eine Sortierung nach Häufigkeit. Die fünf häufigsten Vokale werden alle in 60 Prozent oder mehr der erfassten Sprachen verwendet, und nun ebenfalls in Lugamun (Abb. 2).

PHOIBLE enthält keine Informationen zu Diphthongen, diese sind jedoch in einer anderen Datenbank namens LAPSyD zu finden (Maddieson et al. 2016). Die in den von LAPSyD erfassten Sprachen häufigsten Diphthonge sind **ai** /aᵢ/ ähnlich wie in *Seil*, **au** /aᵤ/ ähnlich *Haus* und **oi** /oᵢ/ ähnlich wie in *Eule*. Wir akzeptieren diese Diphthonge auch in Lugamun, verzichten jedoch auf weitere, die seltener sind (der nächsthäufige wäre /eᵢ/, ähnlich wie in den aus dem Englischen kommenden Wörtern *Mail* und *Fake*).

Was die Anzahl der Konsonanten betrifft, ist laut WALs (Kap. 1) der abgerundete Durchschnitt 22, der Median 21. Lugamun enthält dementsprechend im Prinzip die 22 Konsonanten, die in den Sprachen der Welt am häufigsten vorkommen (wiederum nach PHOIBLE). Vier dieser Konsonanten gelten allerdings als »optional«, d. h., sie können, müssen aber nicht als separate Phoneme ausgesprochen werden. Als Beispiel sei der Glottisschlag [ʔ] genannt, der im Deutschen typischerweise vor Vokalen am Wortanfang sowie vor mit einem Vokal beginnenden Bestandteilen zusammengesetzter Wörter verwendet wird, etwa zwischen den Wortbestandteilen *Spiegel* und *Ei* in *Spiegelei* [ʃpiːgəlʔaᵢ]. Dieser wird (auch) in Lugamun nicht geschrieben, kann jedoch zwischen nebeneinander stehenden Vokalen (etwa in *kruel* [kruʔel] oder [kruːel] – grausam) verwendet werden. Eine Aussprache ohne Glottisschlag, jedoch mit klar voneinander getrennten Vokallauten, ist ebenfalls zulässig, weshalb dieser Konsonant als optional gilt.

	labial	alveolar	postalv.	palatal	velar	glottal
Nasale	m	n				
Plosive	p b	t d			k g	
Frikative	f	s	ʃ (x)			h
Affrikate			tʃ (c) dʒ (j)			
R-Laut		r (r)				
Approximanten		l		j (y)	w (v)	

Abbildung 3: Die Standard-Konsonanten von Lugamun

Nach Abzug der vier optionalen Konsonanten verblieben 18 weitere, die fester Bestandteil von Lugamuns Phonologie sind und in Abb. 3 zu sehen sind. Sofern die verwendete Schreibweise von dem IPA-Symbol abweicht, wird sie in Klammern gezeigt. Der stimmlose postalveolare Frikativ [ʃ] – im Deutschen meist *sch* – wird **x** geschrieben, wie oft im Portugiesischen und in anderen romanischen Sprachen. Der entsprechende Affrikat [tʃ] – im Deutschen meist *tsch* – wird **c** geschrieben, wie im Indonesischen. Das stimmhafte Gegenstück [dʒ] – im Deutschen nur in wenigen Wörtern wie *Dschungel* anzutreffen – wird wie im Englischen als **j** geschrieben. Die Halbvokale [j] und [w] werden als **y** (wie im Englischen und vielen anderen Sprachen) und **v** (wie im klassischen Latein) geschrieben. Das **r** wird vorzugsweise als Flap [ɾ] wie im spanischen *caro* ausgesprochen, da unter den verschiedenen R-Lauten dieser vergleichsweise häufig und auch vergleichsweise einfach zu erlernen ist. Möglich ist aber auch eine Aussprache als »rollendes R« [r] wie im spanischen *perro* oder als Approximant [ɹ] wie im Englischen.

Bei der Auswahl der verwendeten Konsonanten wurden neben deren reiner (in PHOIBLE dokumentierter) Häufigkeit auch in WALS erfasste Regelmäßigkeiten zu deren typischer Verteilung berücksichtigt. So gibt es laut dessen Kapitel 4 nur in einer Minderheit von Sprachen einen Kontrast zwischen stimmlosen und stimmhaften Frikativen. Dementsprechend hat Lugamun zwar ein stimmloses [s] wie im deutschen *dass*, aber kein stimmhaftes [z] wie in *Rose* (es ist aber zulässig, Lugamuns **s** stattdessen als stimmhaftes [z] auszusprechen). Ebenso hat es ein stimmloses [f], aber kein stimmhaftes [v] (der **v** geschriebene Buchstabe kann aber statt als Halbvokal [w] auch als Frikativ [v] ausgesprochen werden). In beiden Fällen wurde der stimmlose anstelle des stimmhaften Konsonanten als regulärer Bestandteil von Lugamuns Phonologie gewählt, da er PHOIBLE zufolge häufiger ist.

Wie man sehen kann, ist Lugamuns Schreibweise ganz regelmäßig; jeder Laut wird immer auf dieselbe Weise geschrieben. Zudem folgt sie dem Prinzip »ein Laut, ein Buchstabe«: jedes Phonem wird durch einen einzelnen Buchstaben repräsentiert; Digraphen werden nur für die Diphthonge verwendet. Anders als etwa im Deutschen oder in Esperanto werden keine diakritischen Zeichen verwendet; die drei Buchstaben *q*, *w* und *z* kommen in Lugamuns Alphabet ebenfalls nicht vor. Betont wird immer der letzte Vokal vor dem letzten Konsonanten (z.B. *ganso*, *etiket*, *sereal*).

Auch die Silbenstruktur orientiert sich an WALS. Dort wird die am meisten verbreitete Silbenstruktur als »mäßig komplex« (*moderately complex*) bezeichnet (WALS 12). Gemeint sind damit Silben der Struktur (C)(L)V(C), wobei C für einen Konsonanten und V für einen Vokal steht. L steht für einen Fließlaut ([l] und [r~r]) oder Halbvokal ([j] und [w]). Kern jeder Silbe ist also ein Vokal (der auch ein Diphthong sein kann). Davor steht optional ein Konsonant – gegebenenfalls auch zwei, aber nur wenn der zweite ein Fließlaut oder Halbvokal ist. Am Silbenende ist ein weiterer Konsonant erlaubt. Das deutsche Wort *Brot* entspricht dieser Silbenstruktur, *Pferd* aber nicht – denn *f* gehört nicht zu den an zweiter Stelle zulässigen Konsonanten und Kombinationen mehrerer Konsonanten am Silbenende sind nicht erlaubt. Alle Wörter in Lugamuns Vokabular bestehen aus einer oder mehreren Silben, die dieser »mäßig komplexen« Struktur entsprechen.

Aber sollten *beliebige* Konsonanten am Silbenende erlaubt sein? Das könnte zu Schwierigkeiten bei der Aussprache führen, da etwa ein [h] am Silbenende für viele ungewohnter und schwieriger auszusprechen ist als eines am Silbenanfang und da beispielsweise Deutsche stimmhafte Verschlusslaute am Silbenende tendenziell als stimmlos aussprechen (aus [d] wird [t]). Manche Sprachen (wie Englisch und Hindi) erlauben sehr viele unterschiedliche Konsonanten am Silbenende, andere sind extrem restriktiv (Japanisch und Mandarin erlauben etwa nur einen bzw. zwei Nasale). Lugamun liegt auch hier eher in der Mitte: in dieser Position sind sechs Konsonanten erlaubt, nämlich **l, m, n, r, s, t**. Diese Auswahl basiert auf einer von mir durchgeführten kleinen Studie zur Häufigkeit unterschiedlicher Konsonanten in dieser Position in Lugamuns Quellsprachen. Die ausgewählten Konsonanten kommen alle in mindestens der Hälfte der Quellsprachen in dieser Position vor (und zwar nicht nur in seltenen Ausnahmefällen), während der Rest seltener ist.

Ebenso wurden auch die am Silbenanfang erlaubten Konsonantenkombinationen noch weiter eingeschränkt, nämlich auf die folgenden:

- *bl, fl, gl, kl, pl, sl*
- *br, dr, fr, gr, kr, pr, tr*
- *cv, dv, gv, hv, kv, sv, tv, xv* (zur Erinnerung: **v** steht für den Halbvokal [w])
- *by, cy, fy, ky, my, ny, py, xy* (**y** steht für den Halbvokal [j])

4 Methodik der Wortauswahl

Bei einer primär auf europäischen Sprachen aufbauenden Plansprache ist es häufig relativ klar, welches für ein Konzept in den Quellsprachen verwendete Wort am besten für die Plansprache geeignet ist, weil es in mehreren der Quellsprachen in ähnlicher Form vorkommt. Bei einer Worldlang ist die Sache hingegen deutlich kniffliger. Nicht selten werden in den von Lugamun berücksichtigten sechs Sprachfamilien komplett unterschiedliche Wörter für ein Konzept verwendet, manchmal sind sogar die Wörter in allen Quellsprachen unterschiedlich. Welches davon soll dann in Lugamun aufgenommen werden, und warum?

Lugamun setzt hierfür auf ein wohldefiniertes Verfahren, das drei Faktoren verbindet:

1. Die Internationalität – ein Wort, das möglichst viele der Quellsprachen gemeinsam haben (bzw. das den in anderen Quellsprachen verwandten Wörtern zumindest ähnlich ist), ist besser als eines, das nur in einer oder zwei vorkommt.
2. Die Originaltreue – ein Wort, das so klingt wie sein Gegenstück in der Quellsprache, ist besser als eines, das stark verändert werden muss, um in Lugamuns Phonologie zu passen.
3. Der Einfluss der Quellsprache – ein Wort, das aus einer bislang in Lugamun unterrepräsentierten Quellsprache kommt, ist besser als eines aus einer der Sprachen, die Lugamuns Wortschatz sowieso schon stark dominieren.

Der letzte Faktor soll dabei versuchen sicherzustellen, dass Lugamuns Wortschatz nicht allzu einseitig von einzelnen Quellsprachen dominiert wird. Insbesondere die beiden romanischen Sprachen sowie das Englische teilen sehr viele Wörter, insbesondere solche, die ursprünglich aus dem Lateinischen kommen. Wir werden später sehen, dass sie deshalb tatsächlich den größten Einfluss auf Lugamuns Wortschatz haben. Zum Ausgleich dafür wird aber in Fällen, in denen die Quellsprachen unterschiedliche Wörter haben, stattdessen eher ein Wort aus einer der in Lugamun schlechter repräsentierten Sprachen verwendet.

Eine weitere Kernidee in Lugamuns Design ist, dass die Auswahl jedes Worts plausibel nachvollziehbar sein muss. Bei natürlichen Sprachen lassen sich über die Herkunft der einzelnen Wörter oft nur mehr oder weniger gut begründete Vermutungen anstellen. Dasselbe gilt für Plansprachen wie Esperanto, Occidental oder Lingua Franca Nova, deren Vokabular zwar aus natürlichen Sprachen übernommen wurde, aber auf Wegen und aus Gründen, die nicht immer genau nachvollziehbar sind. Bei Lugamun wird hingegen für jedes ins offizielle Wörterbuch aufgenommene Wort in einem Auswahl-Logbuch (englisch *selection log*) festgehalten, warum genau dieses Wort gewählt wurde.¹

4.1 Der Auswahlprozess anhand eines Beispiels erläutert

Wie der Prozess genau funktioniert, soll im Folgenden anhand des in Abb. 4 (S. 112) gezeigten echten (nur leicht gekürzten) Auszugs aus dem Auswahl-Logbuch erläutert werden. Die Entwicklung von Lugamun erfolgt mittels Computerunterstützung – ein von mir geschriebenes Programm liest die Wörter mit der passenden Bedeutung aus dem freien Online-Wörterbuch Wiktionary aus, bringt sie in die phonetische Form, die sie in Lugamun annehmen würden, und sortiert sie anhand der drei oben genannten Kriterien. Wie das für das Beispielwort »Gans« – in Lugamun **ganso** – aussieht, ist in der Abbildung gezeigt.

Die primäre Sortierung erfolgt anhand des Kriteriums der *Internationalität*, genauer gesagt anhand der Anzahl weiterer Quellsprachen, die ein ähnliches (oder auch dasselbe) Wort aufweisen. Der erste Kandidat (aus dem Spanischen – es werden immer die ISO-Sprachkürzel verwendet, hier also *es*) ist mit Kandidatenwörtern in drei weiteren Quellsprachen »verwandt« (was

¹ Das komplette Auswahl-Logbuch ist online unter <https://gitlab.com/ChristianSi/lugamun/-/blob/main/data/selectionlog.txt> zu finden. Am selben Ort findet sich unter <https://gitlab.com/ChristianSi/lugamun/-/blob/main/data/dict.txt> auch das komplette mehrsprachige Wörterbuch. Ein leichter lesbares Wörterverzeichnis, das aber nur Lugamuns Wörter und ihre englischen Übersetzungen zeigt, ist unter <https://www.lugamun.org/en/wordlist> zu finden.

damit gemeint ist, wird gleich erläutert). Die beiden nächsten Kandidaten haben verwandte Wörter in nur zwei weiteren Quellsprachen und werden entsprechend weiter unten einsortiert. Danach kommen die Kandidaten mit nur einem verwandten Wort und schließlich (im Bild nicht mehr zu sehen), diejenigen, die gar kein verwandtes Wort in anderen Quellsprachen haben.

```

++ Candidates with 3 related candidates:
[1] es:ganso (P:1.010 – C:0.000+S:0.010+I:1.000, related
    candidates in 3 languages: hi:hanse, id:ansa, ja:gaco,
    ja:gan)
++ Candidates with 2 related candidates:
[2] id:ansa (P:0.537 – C:0.000+S:0.093+I:0.444, related
    candidates in 2 languages: es:ganso, es:ansar,
    hi:hanse)
[3] hi:hanse (P:0.991 – C:0.200+S:0.124+I:0.667, related
    candidates in 2 languages: es:ganso, id:ansa)
++ Candidates with 1 related candidate:
[4]* ja:gan (P:0.000 – C:0.000+S:0.000+I:0.000, related
    candidates in 1 language: es:ganso)
[5] ja:gaco (P:0.207 – C:0.000+S:0.207+I:0.000, related
    candidates in 1 language: es:ganso)
[6] ru:gus (P:0.323 – C:0.000+S:0.100+I:0.222, 1 identical
    candidate, related candidates in 1 language: en:gus)
...
Candidate #1 "ganso" added to the dictionary on 2022-12-15
22:02:13.

```

Abbildung 4: Die Auswahl des Wortes **ganso** (Gans)

Zu beachten ist hierbei, dass diese *relatedness*, auch wenn ich auf Deutsch der sprachlichen Einfachheit halber oft von »Verwandtschaft« rede, tatsächlich die rein *textliche Ähnlichkeit* misst. Zwei Kandidaten gelten als verwandt, wenn sie 50 Prozent ihrer Buchstaben gemeinsam haben, und zwar in der gleichen Reihenfolge. Im gezeigten Beispiel sind die ersten drei Kandidaten tatsächlich etymologisch verwandt: spanisch *ganso*, hindi हंस (*hans*) und indonesisch *angsa* (das »Schwan« oder »Gans« bedeuten kann) sind alle auf dasselbe proto-indoeuropäische Wort zurückzuführen. (In Lugamuns Phonologie wird हंस zu *hanse*, da Sequenzen von zwei Konsonanten am Silbenende nicht erlaubt sind; *angsa*, gesprochen ['aŋsa], wird zu *ansa*, da Lugamun kein Phonem [ŋ] wie in *Ring* kennt.) Hingegen sind die japanischen Wörter 雁 (*gan*) und 鵜 (gachō – bzw. in Lugamuns Schreibweise *gaco*) nicht etymologisch verwandt, sondern teilen nur zufällig die Sequenzen *gan* bzw. *ga...o* mit dem Wort *ganso*. Damit stimmen aber drei von fünf Phonemen bzw. Buchstaben überein, und das reicht für den Algorithmus (der die Etymologie der Wörter ja gar nicht kennt) als Kriterium für hinreichende Ähnlichkeit bzw. »Verwandtschaft«.

Die Sortierung der Kandidatenwörter erfolgt also zunächst nach der Anzahl der ähnlichen Wörter in anderen Quellsprachen, sprich der Internationalität. Im gezeigten Beispiel reicht dies, um *ganso* ganz nach oben zu sortieren, da es der einzige Kandidat mit ähnlichen Wörtern in drei anderen Sprachen ist. Bei gleicher Anzahl ähnlicher Kandidaten wird eine Gesamtpunktzahl (*P* genannt) zur weiteren Sortierung verwendet, die auf den drei genannten Faktoren Originalgetreue (*C* genannt), Internationalität bzw. Ähnlichkeit zu den anderen Kandidaten (*S* genannt) sowie Einfluss der Quellsprache (*I* genannt) basiert. Die errechneten Faktoren werden für jedes Kandidatenwort im Auswahl-Logbuch erfasst und sind entsprechend in Abb. 4 zu se-

hen. Die Einzelfaktoren haben jeweils Werte zwischen 0 (am besten) und 1 (am schlechtesten). Die Gesamtpunktzahl P ist die Summe dieser einzelnen Punktzahlen und kann somit maximal 3 betragen; der minimale und beste Wert ist auch hier 0. Wie in der Abbildung zu sehen, liegt etwa der zweite Kandidat mit einem P -Wert von 0,537 vor dem dritten mit $P = 0,991$.

Wie werden nun die einzelnen Faktoren berechnet, die in P eingehen? Der Faktor C , der die Originalgetreue misst, steigt in Schritten von 0,2 – er kann also 0,0, 0,2, 0,4 usw. sein, aber keine Werte dazwischen annehmen. Für jede größere Veränderung an dem Originalwort, die nötig ist, um es an Lugamuns Phonologie anzupassen, steigt der Wert um 0,2 an. Als größere Veränderung zählt dabei etwa die Einführung oder Löschung eines kompletten Phonems. So wird Hindi हंस (*hans*) wie erwähnt zu *hanse*, da Lugamun keine doppelten Konsonanten am Silbende erlaubt, weshalb in solchen Fällen ein e (als vergleichsweise neutraler, »mittlerer« Vokal) angehängt wird. Für dieses zusätzliche Phonem wird das Kandidatenwort »abgestraft« und erhält einen C -Wert von 0,2.

Der indonesische Kandidat wird zwar auch geringfügig verändert, indem aus *angsa* [ʼaŋsa] *ansa* wird, also [ŋ] zu [n]. Diese Verzerrung gilt aber als hinreichend klein, dass kein Strafpunkt vergeben wird; entsprechend ist der C -Wert 0,0. Als Faustregel gilt hier, dass ein Phonem in ein verwandtes Phonem überführt werden darf, sofern ihre jeweiligen IPA-Symbole von demselben Buchstaben abgeleitet wurden – [ŋ] und [n] sind beide von n abgeleitet, deshalb wird diese Veränderung ohne Strafpunkt akzeptiert. Dasselbe gilt, wenn ein beliebiger R-Laut, wie das spanische »rollende R« [r] oder der im Französischen (und auch im Deutschen) verbreitete stimmhafte uvulare Frikativ [ʁ], in das in Lugamun bevorzugte [r] umgewandelt wird. Da in IPA alle typischen R-Laute bzw. »rhotischen« Konsonanten von r abgeleitet sind, ist auch hier die Anpassung ohne Strafpunkt möglich.

Andere in der Abbildung sichtbare Kandidaten, wie das spanische *ganso* und das japanische 雁 (*gan*), entsprechen Lugamuns Phonologie sogar (nahezu) perfekt; entsprechend erhalten auch sie 0,0 als C -Wert.

Der zweite Faktor ist erneut die Ähnlichkeit zu in anderen Quellsprachen verwendeten Wörtern als Anzeichen für die Internationalität des Wortes. Diesmal wird aber nicht einfach geschaut, welche anderen Quellsprachen ein »hinreichend ähnliches« Wort haben (wie oben beschrieben), sondern per Levenshtein-Distanz gemessen, wie ähnlich ein Kandidatenwort den Kandidaten aus anderen Quellsprachen ist (Levenshtein 1965). Verglichen wird dabei jeweils die in Lugamun verwendete Schreibweise. Hierbei gilt: Je niedriger der Wert, desto ähnlicher sind sich die Wörter, also desto besser. Sind zwei Wörter identisch, ist ihre Levenshtein-Distanz $D = 0$. Ansonsten steigt sie für jeden Buchstaben, der geändert, eingefügt oder gelöscht werden muss, um 1 an. So beträgt die Distanz zwischen *ganso* und *ansa* 2 (»g« wird gelöscht, »o« wird zu »a« geändert).

Die maximale Levenshtein-Distanz zwischen zwei Wörtern entspricht der Länge des längeren Wortes (wenn alle Buchstaben anders sind und somit geändert oder gelöscht werden müssen). Somit bevorzugt diese Metrik inhärent kürzere Wörter – sind zwei kurze Wörter komplett unterschiedlich, ist ihr D -Wert besser als der von zwei komplett unterschiedlichen längeren Wörtern. Um dies auszugleichen, normalisieren wir die Distanz, indem wir sie durch den Durch-

schnitt der Längen der beiden verglichenen Wörter teilen (so dass, wenn beide Wörter gleich lang sind, die maximale normalisierte Distanz $D_n = 1$ beträgt). Die normalisierte Distanz zwischen *ganso* und *ansa* ist also $2 / 4,5 = 0,444$.

Der *S*-Wert (wie in Abb. 4 zu sehen), ist dann die Summe der normalisierten Levenshtein-Distanzen jedes Worts mit den anderen Kandidatenwörtern, erneut normalisiert auf einen Wert zwischen 0 und 1. (Wenn x der beste, d. h. niedrigste Summenwert ist und $x + y$ der höchste, dann wird jede Summe s gemäß der Formel $s_n = (s - x) / y$ umgerechnet.) Im Falle, dass eine Sprache mehrere Kandidatenwörter hat (wie im Beispiel *gan* und *gaco* als japanische Kandidaten), wird für die Distanzberechnung der Wörter in anderen Sprachen jeweils der bessere Wert verwendet.

Im Endergebnis kann man also sagen, dass das Wort mit $S = 0$ den Wörtern in den anderen Quellsprachen »am ähnlichsten« ist, das mit $S = 1$ »am wenigsten ähnlich«.

Das dritte verwendete Kriterium ist der Einfluss der Quellsprache. Dieser Faktor basiert auf der Annahme, dass alle Quellsprachen im Idealfall etwa gleichen Einfluss auf Lugamuns Vokabular nehmen sollten. Ist eine Sprache derzeit also unterrepräsentiert, sollten vorzugsweise Wörter aus dieser Sprache hinzugefügt werden, um das auszugleichen. Dazu ist der *I*-Wert da. Auch dieser Faktor kann Werte zwischen 0 (am besten) und 1 (am schlechtesten) annehmen. Dafür werden die Quellsprachen jeweils vor der Hinzufügung eines neuen Wortes gemäß ihres derzeitigen Einflusses auf das Vokabular sortiert, und zwar in aufsteigender Reihenfolge. Die Sprachen werden dann von 0 anfangend durchgezählt und der jeder Sprache zugeordnete Wert wird durch den höchsten vergebenen Wert (immer 9, da wir zehn Quellsprachen haben) geteilt, so dass alle Ergebnisse im Bereich von 0 bis 1 liegen. Somit erhält die Sprache mit dem geringsten Einfluss den *I*-Wert 0,0, die nächste 0,111, dann 0,222 und so weiter. Die Sprache mit dem höchsten Einfluss erhält somit auch den höchsten (und damit schlechtesten) Wert 1,0.

Als das genannte Beispielwort hinzugefügt wurde, war Spanisch die Sprache mit dem höchsten Einfluss und somit mit $I = 1,0$ (siehe Abb. 4). Dementsprechend erhält sie auch eine ziemlich schlechte Gesamtpunktzahl ($P = 1,01$), die fast völlig von dem *I*-Wert abhängt, da das Wort unverzerrt ist ($C = 0,0$) und den Wörtern in anderen Quellsprachen auch generell sehr ähnlich ist ($S = 0,01$). Da das spanische Wort aber das einzige mit drei verwandten Wörtern in anderen Sprachen ist, kommt die Sortierung nach dem *P*-Wert hier gar nicht zum Zuge und das Wort kann sich trotzdem an die Spitze der Kandidatenliste setzen. Bei den nächsten beiden Kandidaten, *ansa* und *hanse*, sieht man hingegen, dass das indonesische Wort den besseren *I*-Wert hat (0,444 im Vergleich zu 0,667), weil Indonesisch in dem Moment einen geringeren Einfluss hatte als Hindi. Das, zusammen mit der Tatsache, dass der Hindi-Kandidat *hanse* als verzerrt gilt ($C = 0,2$), der indonesische hingegen nicht, hat dann auch maßgeblichen Einfluss darauf, dass *ansa* vor *hanse* gerankt wird.

Wie aber wird der Einfluss der Quellsprachen gemessen? Wenn jedes akzeptierte Wort jeweils nur einer Quellsprache entstammen würde, könnte man einfach die Wörter aus den unterschiedlichen Quellsprachen zählen und so die Reihenfolge der Quellsprachen feststellen. Doch favorisiert der hier verwendete Algorithmus ja gerade besonders »internationale« Wörter, die in möglichst vielen der Quellsprachen zumindest verwandte oder ähnlich aussehende Gegen-

stücke haben. Dass im Beispiel das spanische Wort gewinnt, liegt ganz entscheidend an seiner Ähnlichkeit zu den drei nächsten Kandidaten. Dementsprechend wird, immer wenn ein Wort neu aufgenommen wird, nicht nur die direkte Quellsprache als Einfluss gezählt, sondern auch alle Sprachen, die ein ähnliches Wort haben. (Im Beispiel also neben Spanisch auch Hindi, Indonesisch und Japanisch.) Um den Einfluss der Quellsprachen zu ermitteln, wird dieses Wort also zu je 25 Prozent für jede der vier »Quellsprachen« gezählt.

Bei der Dokumentation neuer Wörter werden auch alle diese Einflüsse erfasst. Zum Beispiel heißt es etwa:

ganso – Gans

De espani *ganso*, hindi हंस (hans) va indonesi *angsa*, va similar a nihon 鷺鳥, 雁 (gachō, gan).

Ins Deutsche übersetzt: »Von spanisch *ganso*, hindi हंस (hans) und indonesisch *angsa*, sowie ähnlich zu japanisch 鷺鳥, 雁 (gachō, gan).«

4.2 Das letzte Wort hat ein Mensch

Der verwendete Algorithmus bringt zwar alle Kandidatenwörter in eine Reihenfolge und schlägt das am besten gerankte Wort für die Auswahl vor. Dieser Vorschlag ist aber nicht zwingend. Es gibt immer noch einen Menschen (im Regelfall der Autor dieses Textes), der das Programm zur Wortauswahl aufruft, die Vorschläge überprüft und die finale Entscheidung über das zu verwendende Wort trifft. In den meisten Fällen ist das tatsächlich der vom Algorithmus vorgeschlagene beste Kandidat, doch keineswegs immer. Es kann eines der auf den weiteren Plätzen gelandeten Wörter sein, oder eine Mischform, die mit keinem der automatisch generierten Kandidaten völlig identisch ist. Die einzige unumstößliche Regel hierbei ist: Die Auswahl muss schriftlich begründet werden, wenn sie von dem Vorschlag des Algorithmus abweicht. Diese Begründung wird zusammen mit dem algorithmischen Ranking im Auswahl-Logbuch festgehalten und macht es so auch in solchen Fällen nachvollziehbar, warum der bevorzugte Kandidat gewonnen hat.

Abb. 5 zeigt dies exemplarisch anhand der Auswahl des Worts **etiket** (Etikett). In diesem Fall wurde *etiketka*, von dem russischen Wort *этикетка* (*etikétka*), auf den ersten Platz gerankt. Ausgewählt als besser geeignetes Wort für Lugamun wurde aber die Form *etiket*, die Indonesisch (Kandidat Nr. 2) und Französisch (Nr. 5 – im Original *étiquette* geschrieben) gemeinsam haben – worauf der Algorithmus durch den Kommentar »1 identical candidate« hinweist. Schaut man sich die vergebenen Gesamtpunktzahlen an, sieht man, dass das russische Wort vor allem deshalb so hoch gerankt wurde, weil Russisch im Moment der Auswahl einen sehr geringen Einfluss auf Lugamuns Vokabular hatte ($I = 0,111$ bedeutet, dass nur eine andere Sprache noch weniger Einfluss hatte). Wie in der gespeicherten Auswahlbegründung (*selection rationale*) festgehalten, wurde hier dem indonesisch/französischen Wort der Vorzug gegeben, weil es kürzer ist (durch den Verzicht auf die rein russische Nachsilbe *-ka*) und von zwei Quellsprachen geteilt wird.

```

++ Candidates with 3 related candidates:
[1]* ru:etiketka (P:0.254 – C:0.000+S:0.142+I:0.111,
  related candidates in 3 languages: es:etiketa,
  fr:etiket, id:etiket)
[2] id:etiket (P:0.433 – C:0.000+S:0.099+I:0.333, 1
  identical candidate, related candidates in 3
  languages: es:etiketa, fr:etiket, ru:etiketka)
[3] id:label (P:0.473 – C:0.000+S:0.140+I:0.333, related
  candidates in 3 languages: en:lebel, hi:lebal,
  ja:raberu)
[4] hi:lebal (P:0.838 – C:0.000+S:0.171+I:0.667, related
  candidates in 3 languages: en:lebel, id:label,
  sw:lebo)
[5] fr:etiket (P:0.877 – C:0.000+S:0.099+I:0.778, 1
  identical candidate, related candidates in 3
  languages: es:etiketa, id:etiket, ru:etiketka)
[6] es:etiketa (P:1.000 – C:0.000+S:0.000+I:1.000,
  related candidates in 3 languages: fr:etiket,
  id:etiket, ru:etiketka)
...
Candidate #2 "etiket" added to the dictionary on 2022-09-
10 15:34:47.
Selection rationale: We prefer this candidate since it's
  shared by two languages and one syllable shorter than
  #1, while being otherwise identical to it.

```

Abbildung 5: Manuelle Korrektur der Auswahl: lieber **etiket** statt **etiketka**.

4.3 Die Wortauswahl und die »großen fünf« Quellsprachen

Der Faktor *I*, der den derzeitigen Einfluss der Quellsprachen berücksichtigt, sorgt für einen gewissen Ausgleich zwischen den Quellsprachen – im Zweifelsfall, wenn nicht andere Kriterien dagegen sprechen, soll so der Einfluss einer derzeit unterrepräsentierten Sprache erhöht werden. Allerdings besteht hier ein gewisses Spannungsverhältnis zu dem Ziel von Lugamun, Wörter zu verwenden, die vielen Menschen schon bekannt vorkommen.

Denn wenn man sich Statistiken der Sprachen mit den meisten Sprechenden anschaut und dabei die verschiedenen Varianten des Arabischen zusammenzählt,² fällt auf, dass es eine große Lücke zwischen den fünf am meisten gesprochenen Sprachen sowie allen anderen gibt. Die »großen fünf« – Englisch, Mandarin, Hindi/Urdu, Arabisch und Spanisch – haben alle mehr als 500 Millionen Sprechende, während alle anderen nur von bis zu 280 Millionen verstanden werden.

Für Lugamun ist das Ziel, prinzipiell Wörter zu wählen, die vielen Sprechenden schon bekannt sind bzw. von ihnen leicht erinnert werden, weil sie dem in einer von ihnen gesprochenen Sprache verwendeten Wort ähneln. Grundsätzlich wird deshalb Wörtern, die in einer möglichst großen Zahl von Lugamuns Quellsprachen zu finden sind, der Vorzug gegeben. Was aber, wenn *alle* Quellsprachen unterschiedliche Wörter für ein Konzept verwenden? Für diesen Fall gilt als Sonderregel, dass ein Wort aus einer der »großen fünf« Sprachen zu wählen ist. So wurde das Wort **tvī** (Bein) aus dem Mandarin übernommen (腿 – tuǐ). Ein Wort aus einer der fünf weiteren Quellsprachen muss hingegen mindestens ein verwandtes bzw. ähnliches Wort in einer anderen Sprache haben, um für die Auswahl in Frage zu kommen. So basiert beispielsweise das Wort **to** (dass) auf dem japanischen と (to) und ist mit dem russischen *umo* (što) verwandt.

² Etwa auf der Website des *Ethnologue*: <https://www.ethnologue.com/insights/ethnologue200/>.

Ohne diesen verwandten Kandidaten wäre es nicht für die Auswahl in Frage gekommen und hätte es daher nicht in Lugamuns Wörterbuch geschafft.

Eine weitere, allerdings nur inoffizielle Faustregel ist, dass eine Verwandtschaft zwischen den beiden romanischen Quellsprachen (Spanisch und Französisch) nicht »zählt«, so sie nicht durch Verwandtschafts- bzw. Ähnlichkeitsbeziehungen mit weiteren Sprachen (meistens dem Englischen) bestätigt wird. Ein rein spanisch/französischer Kandidat wird also so behandelt, als ob er nur *einer* Sprache entstammt, und im Zweifelsfall wird einem Wort, das von zwei anderen Sprachen (in ähnlicher Form) geteilt wird, der Vorzug gegeben, auch wenn dieses eine schlechtere Gesamtpunktzahl aufweist. Dies soll ein Stück weit der Anomalie entgegenwirken, dass die romanischen Sprachen als einzige Sprachfamilie in Lugamuns Quellsprachen doppelt vertreten sind – ansonsten ist jede Sprachfamilie bzw. jeder Zweig des Indoeuropäischen durch maximal eine Sprache repräsentiert.

4.4 Einige Beispielwörter

reforme – reform (verb+noun)

From French *réformer, réforme*, English *reform*, Spanish *reformular, reforma*, Hindi रिफार्म (rifārm), and Russian реформировать, реформа (reformirovat', refórma).

sali – pray

From Swahili *-sali* and Arabic صَلَّى (ṣallā).

sereal – cereal (type of grass and its grains), grain

From Indonesian *sereal*, English/Spanish *cereal*, and French *céréale*.

sore – evening

From Indonesian *sore* and loosely related to French *soir* and Japanese 夜 (yoru).

Derivative: **sore hau** – good evening (greeting)

sosa – suck

From Russian сосать (sosát') and loosely related to Arabic مَضَّ (maṣṣa), Chinese 咂 (za), French *sucer*, and Swahili *-fyonza*.

tenko – weather

From Japanese 天候 (tenkō) and loosely related to Chinese 天氣/天气 (tiānqì) and Spanish *tiempo*.

Abbildung 6: Einige beispielhafte Wörter und ihre Herkunft

Alle Erweiterungen von Lugamuns Wortschatz werden alle paar Wochen auf der Community-Website Reddit dokumentiert, inklusive der Quellsprachen, aus denen jedes Wort hervorgeht. Abb. 6 zeigt exemplarisch einen Ausschnitt aus einem dieser Artikel vom November 2022 (Siefkes 2022), der einige für Lugamun mehr oder weniger typische Kombinationen von Quellsprachen enthält. Das Wort **reforme** (reformieren, Reform) wird von allen indoeuropäischen Sprachen in ähnlicher Form geteilt, taucht aber in keiner anderen Quellsprache auf. **Sali** (beten) kommt aus dem Swahili, das dieses Wort wiederum aus dem Arabischen übernommen hat. **Sereal** (Getreide) wird im Englischen und den romanischen Sprachen, aber auch im Indonesischen verwendet. **Sore** (Abend) ist ein indonesisches Wort, das aber – rein zufällig – auch den französischen und japanischen Wörtern für dieses Konzept ähnelt. **Sosa** (saugen) kommt aus

dem Russischen. Es ähnelt dem französischen Wort, das aus derselben proto-indoeuropäischen Form hervorgegangen ist, aber (vermutlich zufällig) auch den im Arabischen, Chinesischen und Swahili verwendeten Wörtern. **Tenko** (Wetter) kommt aus dem Japanischen. Es ist dem chinesischen Wort für dieses Konzept teilweise verwandt (die erste Silbe) und ähnelt rein zufällig auch dem spanischen Wort.

Diese Beispiele zeigen schön, wie sich Lugamuns Algorithmus sowohl echte, etymologische Verwandtschaften als auch rein zufällige Ähnlichkeiten (»false cognates«) beim Aufbau eines international gut wiedererkennbaren Wortschatzes zunutze macht.

5 Der Einfluss der Quellsprachen

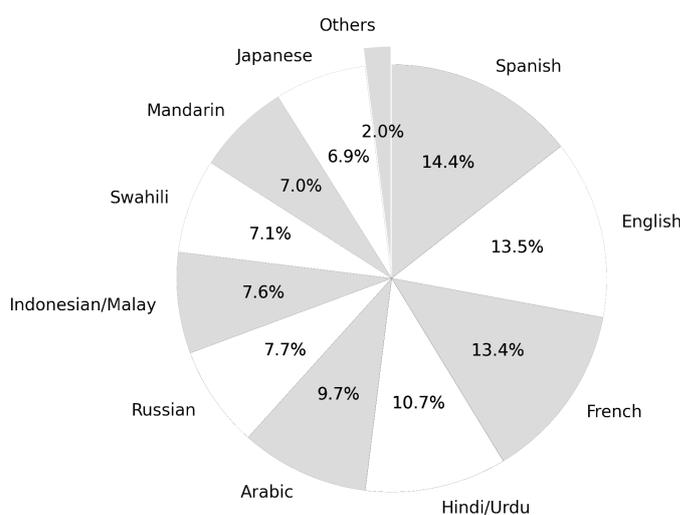


Abbildung 7: Der Einfluss der Quellsprachen

Wie hoch ist nun der Einfluss der verschiedenen Quellsprachen auf Lugamuns Vokabular? Diese Information ist jeweils der Website der Sprache zu entnehmen – Abb. 7 zeigt die Verteilung zum Stand Anfang März 2023.³

Da es zehn Quellsprachen gibt, hätte bei einer perfekten Gleichverteilung jede davon einen Anteil von 10 Prozent. Wie man sieht, weicht die tatsächliche Verteilung stark davon ab. Drei Sprachen – Spanisch, Englisch und Französisch – haben einen Einfluss von über 13 Prozent. Fünf andere – Russisch, Indonesisch, Swahili, Mandarin und Japanisch – kommen jeweils auf unter 8 Prozent. Nur die beiden Sprachen im Mittelfeld – Hindi/Urdu und Arabisch – liegen nahe an dem Ideal von 10 Prozent. (Die zwei Prozent »anderen« erklären sich daraus, dass die für Sprachen, Länder und Volksgruppen verwendeten Wörter generell von dem Endonym abgeleitet werden – so **doice** für »deutsch« und **Lietuva** für »Litauen«.)

Für die Abweichungen von der idealen Verteilung dürfte es primär zwei Ursachen geben: Zum einen die Tatsache, dass die Auswahl von Wörtern zunächst anhand von deren »Internationalität« erfolgt – der Anzahl der Sprachen, die ein ähnliches Wort haben. Und hier punkten vor al-

³ Quelle für Abb. 7–10: <https://www.lugamun.org/en/statistics>; Zugriff jeweils am 2. März 2023. Die Grafiken wurden zur besseren Lesbarkeit farblich nachbearbeitet und nach Schwarzweiß konvertiert.

lem die indoeuropäischen Sprachen und ganz besonders die beiden romanischen Sprachen und das Englische, das viele Wörter von ihnen oder direkt aus dem Lateinischen übernommen hat. Und zum anderen die Bevorzugung der »großen fünf« Quellsprachen, die im Zweifelsfall auch Kandidaten ohne ähnliche Wörter beisteuern können. Vier der fünf am besten repräsentierten Sprachen gehören zu den »großen fünf« – nur Mandarin landet weit abgeschlagen auf dem vorletzten Platz. Hingegen ist es dem Französischen gelungen, den dritten Platz einzunehmen, obwohl es nicht zu den »großen fünf« gehört, also nur aufgrund seiner häufigen Verwandtschaften mit Wörtern in anderen Sprachen.

- **English/Spanish/French**: 36.8%
- Spanish/French/**Russian**: 22.9%
- English/Spanish/**Russian**: 21.9%
- English/French/**Russian**: 21.6%
- Spanish/French/**Indonesian**: 16.7%
- English/French/**Indonesian**: 16.3%
- English/Spanish/**Indonesian**: 16.1%
- Spanish/**Indonesian/Russian**: 14.6%
- English/**Indonesian/Russian**: 14.5%
- French/**Indonesian/Russian**: 14.3%
- Spanish/French/**Hindi**: 13.2%
- English/French/**Hindi**: 12.4%
- English/Spanish/**Hindi**: 12.0%
- English/French/**Japanese**: 12.0%

Abbildung 8: Häufige Kombination von je drei Quellsprachen. Fettdruck markiert jeweils die erste Erwähnung einer Sprache.

Die Verteilung besonders häufiger Kombinationen verwandter bzw. ähnlicher Wörter lässt sich aus weiteren Statistiken ersehen. Abb. 8 zeigt die häufigsten Kombinationen von je drei Sprachen, die zur Auswahl eines Wortes beigetragen haben. Der Prozentsatz drückt dabei aus, bei welchem Anteil von Lugamuns direkt von Quellsprachen abgeleiteten Vokabular diese Kombination im Spiel war.⁴

Wie man sieht, ist die Kombination Englisch/Spanisch/Französisch ganz besonders häufig, da diese Sprachen so viel Vokabular gemeinsam haben. Mehr als ein Drittel von Lugamuns Vokabular basiert auf diesen drei Sprachen (sowie gegebenenfalls weiteren). Dieser hohe Anteil an geteiltem Vokabular erklärt auch, warum sich diese drei Sprachen so deutlich an die Spitze der Einflussverteilung setzen konnten. Die nächsten Kombinationen umfassen je zwei dieser Sprachen gemeinsam mit dem Russischen – eine weitere indoeuropäische Sprache, die mit den anderen europäischen Sprachen ebenfalls viel Vokabular teilt.

Hindi/Urdu ist Lugamuns fünfte indoeuropäische Quellsprache. Daher ist es nicht überraschend, dass es ebenfalls in Kombinationen mit den europäischen Sprachen auftaucht, doch ist sein Anteil zum gemeinsam beigetragenem Vokabular deutlich geringer. Kombinationen zwi-

⁴ Zur Erläuterung: Das oben vorgestellte Wort **ganso** wurde von vier Quellsprachen abgeleitet – jede Kombination von dreien dieser vier Sprachen wird in der Statistik für dieses Wort gezählt.

schen Hindi und dem anglo-romanischen Trio machen nur jeweils etwa 12–13 Prozent von Lugamuns Vokabular aus, während es beim Russischen über 21 Prozent sind.

Kombinationen mit Indonesisch tauchen in der Statistik noch vor denen mit Hindi auf. Obwohl es keine indoeuropäische Sprache ist, hat es einen deutlich höheren Anteil an geteiltem Vokabular mit europäischen Sprachen. Oft handelt es sich dabei um sehr internationale Wörter, die auch noch in weiteren Quellsprachen anzutreffen sind, etwa **dante** (Zahn), **estasion** (Bahnhof), **prinsipe** (Grundsatz, Prinzip), **sukar** (Zucker) und **tradisi** (Tradition).

- **Spanish/French**: 44.7%
- **English/French**: 39.5%
- **English/Spanish**: 39.3%
- **Spanish/Russian**: 25.8%
- **English/Russian**: 24.5%
- **French/Russian**: 24.4%
- **Spanish/Indonesian**: 19.0%
- **French/Indonesian**: 18.7%
- **English/Indonesian**: 18.4%
- **Arabic/Swahili**: 17.1%
- **Indonesian/Russian**: 15.9%
- **English/Japanese**: 15.6%
- **Spanish/Hindi**: 15.6%
- **Arabic/Hindi**: 15.1%
- **French/Hindi**: 15.1%

Abbildung 9: Häufige Kombination von je zwei Quellsprachen

Am Ende taucht im abgebildeten Teil der Statistik noch das Japanische auf, in Kombination mit dem Englischen und dem Französischen. Auch hier handelt es sich meistens um sehr internationale Wörter wie **jangal** (Urwald, Dschungel), **komersia** (Handel, Kommerz) und **tren** (Zug) oder auch um Lehnwörter, die das Japanische meistens aus dem Englischen übernommen hat, etwa bei **leson** (Lektion, Lehrstunde), **memba** (Mitglied, Glied) und **xen** (Kette).

Abb. 9 zeigt die häufigsten Kombinationen von je zwei Quellsprachen in Lugamuns Vokabular. Wenig überraschend dominieren hier Paare aus den bereits betrachteten Dreierkombinationen. Die einzige komplett neue Kombination ist die von Arabisch und Swahili, die immerhin mehr als ein Sechstel von Lugamuns Vokabular ausmacht. Hierbei handelt es sich neben sehr internationalen (und meist aus den indoeuropäischen Sprachen kommenden) Wörtern meist um solche, die Swahili aus dem Arabischen übernommen hat, wie **gali** (teuer), **haki** (Recht, Anspruch), **hata** (sogar) und **lain** (weich).

Die nächsthöchste Kombination zweier nicht-indoeuropäischer Sprachen ist Arabisch mit Hindi. Auch hier handelt es sich oft um Wörter, die – nicht selten über das Persische – aus dem Arabischen übernommen wurden, wie **kanun** (Gesetz, Recht), **hadafe** (Ziel, Zweck) und **suba** (Morgen, Vormittag). Gelegentlich handelt es sich auch um zufällige Ähnlichkeiten, wie bei **jar** (Wurzel), von **جَر** (jar) und ähnlich dem arabischen **جَدْر** (jad̄r).

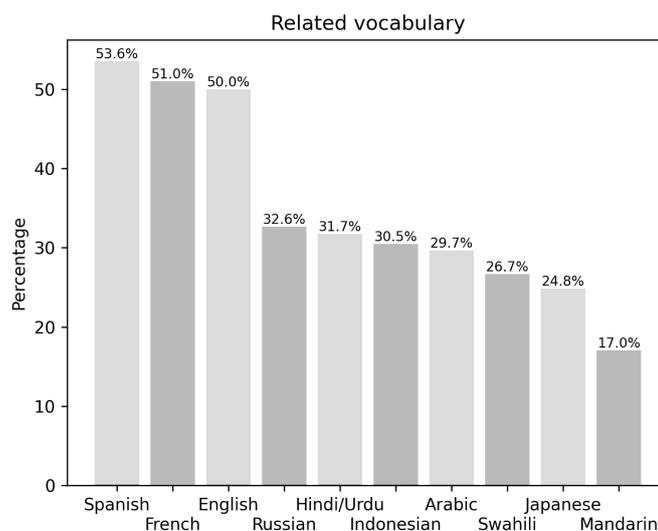


Abbildung 10: Anteil ähnlicher Wörter in den Quellsprachen

Während Abb. 7 den Einfluss der Quellsprachen zeigte, stellt Abb. 10 den Anteil ähnlicher Wörter in den Quellsprachen dar. Wie gesagt wird der Einfluss unter allen Sprachen, deren Wort der gewählten Form zumindest ähnlich ist, gleichmäßig aufgeteilt – er beträgt also insgesamt genau 100 Prozent. Der Anteil ähnlicher Wörter wird hingegen für jeden Fall von Verwandtschaft oder Ähnlichkeit erhöht, unabhängig davon, wie viele andere Sprachen ebenfalls ein ähnliches Wort aufweisen. Wäre jedes Wort nur von genau einer Quellsprache abgeleitet, würde die gesamte Prozentzahl hier also ebenfalls nur 100 betragen, doch da Lugamun viele internationale Wörter aufweist, die in vielen Sprachen zumindest in ähnlicher Form auftreten, ist es nicht überraschend, dass die Gesamtzahl deutlich höher ausfällt.

Hier zeigt sich noch mehr als in der anderen Grafik ein deutlicher Vorteil für die fünf indoeuropäischen Quellsprachen: Jede von diesen weist einen höheren Anteil ähnlichen Vokabulars auf als jede der weiteren Quellsprachen. Da sie so viel Vokabular miteinander (und oft auch mit anderen Sprachen) teilen, führen die beiden romanischen Sprachen und das Englische das Feld deutlich an, mit 50 Prozent oder mehr geteiltem Vokabular. Deutlich abgeschlagen, aber immer noch mit um die 32 bis 33 Prozent ähnlichen Wörtern, folgen Russisch und Hindi/Urdu. Nur geringfügig dahinter, mit um die 30 Prozent, liegen Indonesisch und Arabisch. Dem Indonesischen nutzt dabei insbesondere, dass es viele europäische Wörter übernommen hat. Hingegen profitiert das Arabische davon, als Leihgeber viele Wörter an andere Quellsprachen (zumindest Swahili und Hindi, aber auch Indonesisch) übergeben zu haben.

Etwas abgeschlagen mit etwa 25 bis 27 Prozent folgen Swahili und Japanisch. Geradezu dramatisch sieht die Situation bei Mandarin aus, das mit 17 Prozent das Schlusslicht bildet. Das ist zwar immer noch deutlich mehr, als die theoretischen 10 Prozent bei »gar kein geteiltes Vokabular« betragen würden. Dennoch sieht man hier sehr deutlich, dass die chinesischen Sprachen fast schon »eine Welt für sich« sind, da sie nur äußerst selten Vokabular mit anderen Sprachen teilen. Wo sie es dennoch tun, ist es (unter den Quellsprachen) meistens mit dem Japanischen, das so manches chinesische Wort in ähnlicher Form übernommen hat, etwa **jen** (Mensch) oder **pifu** (Haut).

6 Textbeispiel und einige Hinweise zur Grammatik

Als kurzes Textbeispiel möge die Übersetzung eines Absatzes aus Martin Luther Kings berühmter »I Have a Dream«-Rede dienen:

Mi xvo a tum si den, mis rafiki, malgre to nas konfronta yo difisil de si den va den tali, mi ha yexo sonya. It xi sonya gen jar profon ni sonya emeriki. Mi ha sonya to aru den si nasi-on ga kaixu janli va fa real o manan sahi de iti imanin: "Nas opin to si yo sahi eviden: to ol jen bi krea kom igval." Mi ha sonya to aru den ni yo komonte lal de Georgia, yo bet de esklave prei va bet de haja esklave prei ga bisa side gengen ni mesa de kikanes.

[Ich sage Ihnen heute, meine Freunde, auch wenn wir mit den Schwierigkeiten von heute und morgen konfrontiert sind, so habe ich doch einen Traum. Es ist ein Traum, der tief im amerikanischen Traum verwurzelt ist. Ich träume davon, dass sich diese Nation eines Tages erhebt und die wahre Bedeutung ihres Glaubensbekenntnisses auslebt: »Wir halten diese Wahrheiten für selbstverständlich: dass alle Menschen gleich geschaffen sind.« Ich träume davon, dass eines Tages auf den roten Hügeln Georgias die Söhne ehemaliger Sklaven und die Söhne ehemaliger Sklavenhalter sich gemeinsam an den Tisch der Brüderlichkeit setzen können.]

Kurze Erinnerung zur Aussprache: **x** wird wie das deutsche *sch* ausgesprochen, **c** (kommt im Textbeispiel nicht vor) wie *tsch*, **j** ist sein weiches Gegenstück wie in *Dschungel*; das **s** ist immer »scharf«, also stimmlos. **Y** wird wie deutsches *j* ausgesprochen, **v** vorzugsweise als Halbvokal wie im englischen *wet*; **ai**, **au** und **oi** sind Diphthonge. Betont wird immer der letzte Vokal vor dem letzten Konsonanten: *rafiki*, *malgre*, *difisil*. Gibt es keinen solchen Vokal, wird der erste betont: **krea**.

Zur Grammatik können hier nur einige kurze Erläuterungen anhand des Textbeispiels gebracht werden; eine detaillierte Grammatik ist auf der Website des Projekts zu finden.⁵ Lugamuns Grammatik ist analytisch – Wörter ändern ihre Form fast nie für rein grammatische Zwecke. Verben und Adjektive können generell auch als Substantive verwendet werden, mit der Bedeutung *der Akt X* bzw. *das, was X ist*. So bedeutet **sonya** »träumen« und »Traum«, **iman** »glauben« und »Glaube«, **difisil** sowohl »schwer, schwierig« als auch »Schwierigkeit« (etwas, das schwierig ist).

Adjektive folgen dem Substantiv, das sie modifizieren: **den tali** (nächster Tag, morgen), **sonya emeriki** (amerikanischer Traum). Es gibt jedoch eine kleine Anzahl von Wörtern – »Selektoren« und »Quantifikatoren« genannt – die vor dem Substantiv stehen (die meisten von ihnen können auch alleine als Pronomen verwendet werden). Dazu gehören **yo** (Plural: zwei oder mehr), **ingi** (viele), **ol** (alle), **si** (dieser), **ta** (jener), **aru** (irgendeiner, ohne genauere Angaben welcher). Die Markierung des Plurals ist optional; wird die Mehrheit schon durch ein anderes Wort wie **ingi**, **ol** oder ein Zahlwort ausgedrückt, wird **yo** niemals verwendet. Ansonsten kann darauf verzichtet werden, wenn der Sprecher davon ausgeht, dass sich die Mehrheit aus dem Kontext ergibt. So wird in der Formulierung **mis rafiki** (mein Freund, meine Freunde) darauf verzichtet, da aus der Redesituation klar ist, dass King viele Menschen adressiert.

Affixe werden für die Erweiterung des Wortschatzes verwendet, um verwandte Wörter aus einem Grundwort abzuleiten. So werden Possessive durch das Anhängen eines Lauts aus dem

⁵ <https://www.lugamun.org/en/grammar/intro>

entsprechenden Personalpronomen abgeleitet. Nach einem Vokal wird **-s** angehängt, so wird aus **mi** (ich) **mis** (mein). Nach Konsonanten wird stattdessen **-i** verwendet, entsprechend wird **it** (es) zu **iti** (sein, ihr – bezogen auf Dinge, nicht Personen).

-i ist auch generelles Suffix zur Erzeugung von Adjektiven; ein Vokal am Ende des Grundwortes fällt dabei weg, sofern er nicht der einzige ist. So wird aus **Emerika** (USA) **emeriki** (US-amerikanisch), aus der Präposition **pre** (vor) wird **prei** (ehemalig, früher). **Emerika** wird für das Land im Unterschied zu **Amerika**, den Kontinenten, verwendet.

Es gibt diverse weitere Affixe; zudem können manche Präposition wie etwa **pre** auch als Präfix verwendet werden. Zu den im Text sichtbaren Suffixen gehört **-in**, das das typische Objekt eines Verbs ausdrückt. So wird aus **kula** (essen) **kulin** (Nahrung, Essen – das, was man isst); aus **iman** (glauben, Glaube) wird **imanin** (Glaubensbekenntnis – der Inhalt bzw. Ausdruck dessen, woran man glaubt). **-ja** bezeichnet die Person, die etwas tut, so wird aus **safir** (reisen) **safirja** (Reisender); aus **ha** (haben, besitzen) wird **haja** (Inhaber, Besitzer, Eigentümer). **-nes** erzeugt abstrakte Substantive, oft entsprechend dem deutschen *-heit* oder *-keit*: **huru** (frei) – **hurunes** (Freiheit), **kika** (Geschwister) – **kikanes** (Geschwisterlichkeit)

Personenbezogene Substantive erhalten (von wenigen Ausnahmen abgesehen) keine Angaben über das Geschlecht der Person; so gewünscht, kann eine explizit männliche Form durch das Prefix **ma-**, eine weibliche durch **nu-** markiert werden. So wird aus **kika** **makika** (Bruder) oder **nukika** (Schwester). Die Übersetzung von M. L. Kings Rede ist etwas geschlechtsneutraler formuliert als das Original und seine deutsche Übersetzung, da sie etwa von **yo bet** (Nachkommen) statt **yo mabet** (Söhne) spricht.

Ein weiteres Präfix ist **ko-**, das eine kleinere Variante von etwas ausdrückt: **monte** (Berg) – **komonte** (Hügel). Generell behalten Präfixe die Wortart des Wortes bei, an das sie angehängt werden, während jedes Suffix Wörter einer bestimmten Wortart erzeugt (**-i** erzeugt Adjektive; **-in**, **-ja** und **-nes** Substantive).

Zwei nebeneinander platzierte Substantive werden implizit durch die Genitiv-Präposition **de** verbunden, so bedeutet **den can** »Tag der Geburt«, also »Geburtstag«, ein **haja esklave** ist ein »Inhaber von Sklave[n]«, also ein »Sklavenhalter«. Auch Reduplikation wird zur Erzeugung neuer Wörter verwendet. Kurze Präposition (mit nur einer Silbe) werden durch Verdoppelung zu Adverbien, so wird aus **gen** (mit) **gengen** (zusammen, gemeinsam).

Tempus, Aspekt, Modus und Handlungsrichtung von Verben werden durch vor das Verb gesetzte Partikel ausgedrückt. **Li** markiert die Vergangenheit, **na** die Gegenwart und **ga** die Zukunft. Diese Tempus-Partikel sind optional und werden normalerweise weggelassen, wenn auch so klar ist, wann der bezeichnete Akt stattfindet. **Bi** setzt das Verb ins Passiv: **bi krea** – erschaffen/geschaffen werden. Daneben gibt es zwei Modus-Partikel: **ba** für den Potentialis (entsprechend Esperanto *-us*), **du** für den Imperativ bzw. Adhortativ (*-u* in Esperanto). **Sai** drückt den Progressiv aus (entsprechend dem englischen *-ing*, aber deutlich sparsamer verwendet).

Tu markiert den Infinitiv. Es steht zwischen verketteten Verben, da andernfalls das zweite Verb als Substantiv interpretiert würde: **Mi yau tu dansa** – Ich will tanzen. / **Mi yau dansa** – Ich

will einen Tanz. Verben mit **tu** (und optional einem eigenen Objekt) können aber auch wie Substantive verwendet werden, um sich auf eine Aktivität im Allgemeinen zu beziehen: **Tu dansa tango no asan** – Tango (zu) tanzen ist nicht einfach. Zwischen verketteten Verben kann **tu** zu **-u** verkürzt werden; dann wird es als Suffix an das erste Verb angehängt und ersetzt ggf. dessen finalen Vokal: **kaixi tu janli** bzw. **kaixu janli** bedeutet »beginnen zu stehen«, d. h. »sich erheben, aufstehen«. Es gibt nur zwei Hilfsverben – **bisa** (können) und **debe** (müssen) –, denen immer ein Verb folgt, ohne dass **tu** verwendet werden müsste.

Die übliche Wortstellung ist Subjekt-Verb-Objekt. Optional kann das Subjekt durch die Partikel **i** eingeleitet werden, das Objekt durch **o**. Dies ermöglicht Abweichungen von der normalen Wortstellung, etwa zur Betonung. **O** muss verwendet werden, wenn das Verb auch ein Adjektiv-Komplement hat, um klarzumachen, wo dieses endet und das Objekt beginnt: **fa real o manan sahi** – »die wahre Bedeutung verwirklichen/realisieren« (wörtlich: real machen).

Generell ist Lugamuns Grammatik durch zwei zusammenhängende Leitideen geprägt: möglichst »einfach« zu sein und zwar, soweit dies möglich ist, unabhängig von dem sprachlichen Hintergrund des Lernenden; und möglichst »durchschnittlich« zu sein, in dem Sinne, dass möglichst viele Sprachen ähnliche Merkmale aufweisen. Für letztere Eigenschaft wurde, ebenso wie bei der Phonologie, wiederum intensiv auf in WALS (Dryer & Haspelmath 2013) zu findende Informationen zurückgegriffen; zudem wurde berücksichtigt, wie die zehn Quellsprachen mehrheitlich an bestimmte Problemstellungen herangehen. Dass Adjektive hinter statt vor dem Substantiv stehen, folgt etwa der deutlichen Mehrheit der Sprachen nach WALS (Kap. 87). Hingegen gibt es zwei sehr verbreitete Wortstellungen, mit einem geringen Vorsprung von Subjekt-Objekt-Verb (SOV) vor Subjekt-Verb-Objekt (SVO) (Kap. 81). In diesem Fall wurden die Quellsprachen herangezogen, und da acht von ihnen SVO verwenden, tut Lugamun das ebenfalls.

Wo Einfachheit und Durchschnittlichkeit in Konflikt zueinander zu stehen scheinen, wurde der Einfachheit der Vorzug gegeben. Ein weiteres Kriterium ist, dass die Sprache »natürlich« wirken soll. Deshalb wurde auf strikte Wortartenmarker wie in Esperanto verzichtet, da diese in ihrer Rigorosität in keiner natürlichen Sprache zu finden sind.

7 Die Geschichte und Zukunft von Lugamun

Lugamun ist ein junges Projekt zur Schaffung einer Worldlang mit einer allerdings unsicheren Zukunft. Entscheidend für seine Geschichte sind zwei miteinander verbundene Online-Communitys: das *Auxlangs*-Forum auf der Community-basierten Website Reddit⁶ und der *auxlangs/helpingvoj*-Server der Chatanwendung Discord.⁷ Nach einigen Vorarbeiten habe ich im Frühjahr 2021 in dem Reddit-Forum die Phonologie und Wortauswahl-Strategie des Sprachprojekts vorgestellt (Siefkes 2021a, 2021b) und nach Rückmeldungen von anderen Nutzern weiter verfeinert. Im Juli desselben Jahres war die von mir parallel dazu entwickelte Software für die Wortauswahl dann soweit, dass die eigentliche Arbeit an Lugamuns Vokabular beginnen

⁶ <https://www.reddit.com/r/auxlangs/>

⁷ Um Discord-Servern beizutreten, muss man sich »einladen« lassen. Ein Einladelink für diesen Server ist im zuvor genannten *Auxlangs*-Forum zu finden.

konnte. Das älteste im Auswahl-Logbuch dokumentierte Wort, das bis heute in unveränderter Form überlebt hat, ist passenderweise **un** (eins).

Vor allem für die weitere Arbeit am Vokabular, die Ausarbeitung und Dokumentation der Grammatik sowie das Erstellen diverser kleinerer Beispieltex-te und Übersetzungen waren der Discord-Server und eine kleine Gruppe dort zu findender »Fans« des Sprachprojekts entscheidend. Alle dauerhaften Ergebnisse werden auf der Website des Projekts festgehalten.⁸ Diese ist ein offenes Wiki, bei dem sich alle Interessenten unbürokratisch registrieren können, um daran mitzuwirken. Der längste dort bislang zu findende Text in Lugamun ist eine von einem der ersten Nutzer der Sprache erstellte Übersetzung von Ernest Hemingways Kurzgeschichte »Hills Like White Elephants«.⁹

In Lugamuns Wörterbuch sind derzeit (Stand 7. März 2023) 1600 Einträge zu finden, darunter gut 980 direkt aus den Quellsprachen abgeleitete Wortwurzeln. (Der Rest sind abgeleitete bzw. zusammengesetzte Wörter und einige wenige Phrasen.) Die weitere Zukunft der Sprache ist allerdings unsicher, da ich dabei bin, mich aus dem Projekt zurückzuziehen (siehe Siefkes 2023 für Details und Begründung). Sofern sich nicht jemand anderes findet, um insbesondere den sehr aufwändigen Wortauswahlprozess zu übernehmen, dürfte die Weiterentwicklung der Sprache deshalb bald zum Erliegen kommen. So wird Lugamun vielleicht eher ein »proof of concept« bleiben, wie eine naturalistische, aber dennoch regelmäßige Worldlang aussehen könnte, die sich systematisch an Wissensquellen wie PHOIBLE und WALS orientiert und auf algorithmische Unterstützung für die Auswahl geeigneter internationaler Wörter setzt, als selbst zu einer weit entwickelten, vielseitig einsetzbaren Sprache zu werden.

Literatur

- Dryer, Matthew S. & Haspelmath, Martin (Hrsg.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max-Planck-Institut für evolutionäre Anthropologie. <http://wals.info>.
- Fetcey, Staren & Comité Linguistique Kotava. 2022. *Kotava – Grammaire officielle complète*. Version IV.08: Décembre 2022. http://www.kotava.org/fr/fr_pulviropa_000.pdf.
- IAL Wiki. 2019. *Worldlang*. <https://ial.fandom.com/wiki/Worldlang>.
- Levenshtein, Vladimir I. 1965. *Binary codes capable of correcting deletions, insertions, and reversals*. In *Doklady Akademii Nauk SSSR* 163(4), 845–848 (russisch, englische Übersetzung in *Soviet Physics Doklady*, 10(8), 1966, 707–710).
- Maddieson, Ian & Flavier, Sébastien & Marsico, Egidio & Pellegrino, François, 2016. *LAPSYD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0*. Lyon: Laboratoire Dynamique de Langage. <https://lapsyd.huma-num.fr/lapsyd/>.
- Moran, Steven & McCloy, Daniel (Hrsg.). 2019. *PHOIBLE 2.0*. Jena: Max-Planck-Institut für Menschheitsgeschichte. <http://phoible.org>.
- Siefkes, Christian. 2021a. An “average” phonology and spelling for a worldlang. *Reddit*, 28. März. https://www.reddit.com/r/auxlangs/comments/mf4h7k/an_average_phonology_and_spellin_g_for_a_worldlang/.

⁸ <https://www.lugamun.org/>

⁹ https://www.lugamun.org/trans/yo_komonte_kom_elefan_bai

126 Christian Siefkes: *Lugamun – eine mit algorithmischer Unterstützung erstellte »Worldlang«*

Siefkes, Christian. 2021b. Vocabulary selection for a worldlang. *Reddit*, 6. April. https://www.reddit.com/r/auxlangs/comments/mlf8h8/vocabulary_selection_for_a_worldlang/.

Siefkes, Christian. 2022. New words in Lugamun's dictionary (week 46). *Reddit*, 15. Nov. https://www.reddit.com/r/auxlangs/comments/yvw3k5/new_words_in_lugamuns_dictionary_week_46/.

Siefkes, Christian. 2023. Further Lugamun development will soon be paused. *Reddit*, 18. Feb. https://www.reddit.com/r/auxlangs/comments/115babo/further_lugamun_development_will_soon_be_paused/.

Über die Autoren

Věra Barandovská-Frank (barandov@mail.uni-paderborn.de), Dr. phil., ist Latinistin und Romanistin, unterrichtet Interlinguistik an der Universität Posen.

Pierre Dieumegard (pierre.dieumegard@free.fr) ist Doktor der Pflanzengenetik, emeritierter Lehrer für Biologie und Erdkunde, Vorsitzender der Gesellschaft »Espéranto-Développement-45« in Orléans (Frankreich).

Sabine Fiedler (sfiedler@uni-leipzig.de), Prof. Dr. phil. habil., ist Sprachwissenschaftlerin am Institut für Anglistik der Universität Leipzig. Seit 2011 ist sie Vorsitzende der Gesellschaft für Interlinguistik e. V.

Chloé Izquierdo ist Magisterstudentin der Sprachwissenschaft an der Universität Orléans, 2021–2022 Volontärin bei der Gesellschaft »Espéranto-Développement-45«.

Goro Christoph Kimura (g-kimura@sophia.ac.jp), Dr. phil., ist Professor für deutsche und europäische Studien an der Fakultät für Fremde Sprachen und Studien der Sophia-Universität Tokyo und war 2021–2022 Gastwissenschaftler an der Universität Leipzig.

Cornelia Mannewitz (cornelia.mannewitz@gmx.de), Dr. phil. habil., Slawistin, nach Beendigung ihrer Tätigkeit in universitärer Lehre und Forschung 2021 aktiv in der GIL und in der Friedensbewegung.

Klaus Schubert (ks@trans-kom.eu; www.klausschubert.de), Dr. phil. habil., ist Professor der Universität Hildesheim im Ruhestand. Sein Forschungsgebiet ist die Angewandte Sprachwissenschaft, insbesondere Interlinguistik, Fachkommunikationsforschung und Translationswissenschaft.

Christian Siefkes (christian@siefkes.net), Dr., promovierte im Bereich Maschinelles Lernen und lebt als freiberuflicher Softwareentwickler und Autor in Berlin.