

INFORMATIONSDICHTE UND DIE VORHERSAGBARKEIT DER PHONETISCHEN STRUKTUR

Bistra Andreeva, Bernd Möbius, Ivan Yuen, Omnia Ibrahim
Universität des Saarlandes (Deutschland)

INFORMATION DENSITY AND THE PREDICTABILITY OF PHONETIC STRUCTURE

Bistra Andreeva, Bernd Möbius, Ivan Yuen, Omnia Ibrahim
Saarland University (Germany)

<https://doi.org/10.60055/GerSk.2025.izv.1.252-267>

Abstract: Die Studie untersucht die Beziehung zwischen Informationsdichte und linguistischer Kodierung in der Phonetik sowie der menschlichen Sprachverarbeitung. Die Informationsdichte einer linguistischen Einheit wird in Bezug auf Surprisal (den negativen Logarithmus der Wahrscheinlichkeit einer Einheit in einem gegebenen Kontext) definiert. Die Effekte von Surprisal auf die phonetische Kodierung wurden hinsichtlich verschiedener Aspekte wie Formantenverläufe von Vokalen, Stimmhaftigkeit von Plosiven, Silbendauer und Vokaldispersion (auch im L2) untersucht, wobei Kontrollfaktoren der prosodischen Struktur sowie mögliche Interaktionen mit dem Lombard-Effekt und der prosodischen Struktur berücksichtigt wurden. Die Ergebnisse deuten darauf hin, dass Sprecher phonetische Details anpassen, um ein Gleichgewicht zwischen Informationsdichte und phonetischer Kodierung aufrechtzuerhalten.

Schlüsselwörter: Surprisal, Formante, Lombard-Effekt, Stimmhaftigkeit, L2

Abstract: The study investigates the relation between information density and linguistic encoding in phonetics and human speech processing. Information density of a linguistic unit is defined in terms of surprisal (the unit's negative log probability in a given context). The effects of surprisal on phonetic encoding have been examined with respect to various aspects such as the formant trajectories of vowels, the voicing

of plosives, syllable duration, and vowel dispersion (also in L2), considering control factors of prosodic structure and potential interactions with the Lombard effect and the prosodic structure. The results suggest that speakers adjust phonetic details to maintain a balance between information density and phonetic encoding.

Keywords: surprisal, formants, Lombard effect, voicing, L2

1. Einleitung

Die vorliegende Arbeit befasst sich mit der Beziehung zwischen Informationsdichte und sprachlicher Kodierung in der Produktion der gesprochenen Sprache. Zur Modellierung dieser Beziehung werden in neuerer Zeit informationstheoretische Ansätze verfolgt, die auf Shannons Informationsbegriff zurückgehen (Shannon 1948) und ihn methodisch erweitern. Die Informationsdichte einer linguistischen Einheit (Laut, Silbe, Wort etc.) wird als Surprisal definiert. Der Surprisal quantifiziert wie vorhersagbar diese Einheit in einem gegebenen Kontext ist. Die Haupthypothese, die unserer experimentellen und Modellierungsarbeit zugrunde liegt, ist, dass Sprecher Details der phonetischen Kodierung modulieren, um ein Gleichgewicht der komplementären Beziehung zwischen Informationsdichte und phonetischer Kodierung aufrechtzuerhalten. Um diese Hypothese zu testen, analysieren wir die Auswirkungen des Surprisals auf die Formantverläufe der Vokale, die Segmentdauer, die Größe des Vokalraums und die Stimmhaftigkeit der Plosive, indem wir in unseren statistischen Modellen für die prosodische Struktur (Wortbetonung, prosodische Grenzen und Sprechgeschwindigkeit) kontrollieren.

Darüber hinaus untersuchen wir den Einfluss der prosodischen Struktur (Tonhöhenakzente und Grenzstärke) und der kontextbezogenen Vorhersagbarkeit auf die Dauer der Silbe vor dem Ende der Intonationsphrase. Die Ergebnisse zeigen eine Wechselwirkung zwischen Surprisal und Stärke der prosodischen Grenze. Die Silbendauer wird mit zunehmendem Surprisalwert länger, aber diese Zunahme der Dauer ist systematisch größer vor starken als vor schwachen Grenzen. Diese Modulation der Silbendauer wird über den bereits etablierten Effekt der phrasenfinalen Silbendehnung hinaus beobachtet.

Schließlich untersuchen wir, inwieweit informationstheoretische Faktoren der Zielsprache die phonetischen Ausprägungen der L2-Realisierungen von Sprachlernern erklären können, auch in Bezug auf unterschiedliche Niveaustufen der Sprachkompetenz. Die Ergebnisse deuten darauf hin, dass bulgarische L2-Sprecher des Deutschen mit einem fortgeschrittenen Sprachkompetenzniveau (C2) die Surprisalwerte bei der Realisierung der deutschen Vokale berücksichtigen, indem sie den Vokalraum der unvorhersagbaren Vokale vergrößern. Bulgarische L2-Sprecher des Deutschen mit einem mittleren Sprachkompetenzniveau (B2) wenden bulgarische Reduktionsmuster bei den vorhersagbaren Vokalen an, indem sie den Vokalraum reduzieren und anheben.

2. Methodologie

2.1. Sprachmodellierung

Der silben- und segmentbasierte Surprisal wurde in allen nachfolgend diskutierten Studien mithilfe von Sprachmodellen geschätzt, die auf dem deWaC-Korpus (Baroni et al. 2009) trainiert wurden. Das Korpus umfasst etwa 1,7 Milliarden Worttokens und 8 Millionen Worttypen aus einer Vielzahl von Genres wie Zeitungsartikeln und Chatnachrichten. Es wurde mit Hilfe der deutschen Version des Festival Text-to-Speech Synthese Systems (Möhler et al. 2000) vorverarbeitet und normalisiert, wobei irrelevante und doppelte Dokumentinformationen, wie HTML-Strukturen oder lange Listen, entfernt wurden. Nach der Vorverarbeitung wurde das normalisierte Korpus in eine Trainingsmenge (80 %) und eine Testmenge (20 %) aufgeteilt. Die silbenbasierten Trigramm-Sprachmodelle, die die Wortgrenze als Einheit berücksichtigen, und die segmentbasierten Modelle, die die Silbengrenze als Einheit einbeziehen, wurden mithilfe des SRILM-Toolkits (Stolcke 2002) auf der Trainingsmenge trainiert und anschließend mit der Witten-Bell-Glättung (Witten / Bell 1991) verfeinert. Die Modelle dienten zur Berechnung der bedingten Wahrscheinlichkeiten einer Silbe oder eines Segments im gegebenen Kontext, wobei der Kontext durch die beiden vorhergehenden Einheiten (inklusive Wort- oder Silbengrenze) definiert wurde:

$$S(\text{unit}_i) = -\log_2 P(\text{unit}_i | \text{unit}_{i-1}, \text{unit}_{i-2}),$$

S steht für Surprisal und P steht für die Wahrscheinlichkeit (Hale 2016). Die bedingten Wahrscheinlichkeiten bildeten somit die Surprisalmaße für die Zielsilben und -segmente.

2.2. Sprachdatensätze

Die meisten Studien in dieser Arbeit basieren auf existierenden Korpora, die Sprachdaten enthalten, die von anderen Forschergruppen aus verschiedenen Gründen erhoben und linguistisch annotiert wurden, sowohl automatisch als auch manuell.

Das für die Analyse der Formantverläufe verwendete Sprachkorpus (Abschnitt 3.1.1) ist das Siemens Synthese-Korpus (SI1000P; Schiel 1997), das qualitativ hochwertiges Material für die konkatenative Sprachsynthese bereitstellt. Das Korpus enthält Aufnahmen von zwei professionellen männlichen Sprechern des Standarddeutschen, die 992 Sätze aus dem Zeitungskorpus der Frankfurter Allgemeinen (SI1000) vorlasen und in einem schallisolierten Aufnahmestudio mit einer Abtastrate von 48 kHz und 16 Bit aufgenommen wurden. Diese Aufnahmen wurden gefiltert und auf 16 kHz heruntergesampelt. Kanonische Transkriptionen und automatische Segmentierungen sind verfügbar.

Für die Untersuchung der Stimmhaftigkeit von Plosiven (Abschnitt 3.1.2) sowie der Wirkung von Kanalmerkmalen auf Silben mit unterschiedlichen Sur-

praisalwerten (Abschnitt 3.2) wurde ein Lombard-Korpus deutscher Lesesprache erstellt. Aus dem deWaC-Korpus wurden 60 Sätze auf der Grundlage von Bins mit hohen und niedrigen Surprisalwerten ausgewählt und so gekürzt, dass die Zielsilben und ihre Surprisalwerte erhalten blieben. Jede CV-Zielsilbe enthielt Plosive (/p, b, d, k/) in Kombination mit den fünf langen Vokalen (/a:, e:, i:, o:, u:/). Die Geräuschpegel (0 dB und -10 dB Signal-Rausch-Verhältnis) wurden systematisch variiert, während der mittlere Block als Referenz ohne Vorhandensein von Geräuschen aufgezeichnet wurde. Das Material wurde von 38 deutschen Muttersprachlern (12 Männer, 26 Frauen) produziert. Die Aufnahmen wurden als Mono-WAV-Dateien mit einer Abtastrate von 48 kHz und 24 Bit gespeichert.

Für die Analyse möglicher Wechselwirkungen zwischen Surprisal und prosodischer Struktur auf die phrasenfinale Silbe (Abschnitt 3.3) wurde eine Teilmenge der Diskurs-Informationen-Radio-Nachrichten-Datenbank für Linguistische Analysen (DIRNDL) verwendet (Eckart et al. 2012; Björklund et al. 2014). Dieses Korpus umfasst etwa fünf Stunden gelesener Nachrichten, die von neun professionellen Sprechern (5 Männer, 4 Frauen) produziert und in Wörter, Silben und Phoneme segmentiert wurden. Die Annotation von Tonhöhenakzenten und prosodischen Grenzen erfolgte manuell durch eine studentische Hilfskraft, die das GToBI(S)-System (Mayer 1995) als Grundlage nutzte.

Eine Ausnahme bildet die Studie zur Erhebung der Informationsdichte als potenzielles Maß für Sprachkompetenz (Abschnitt 3.4), in der jeweils sechs bulgarische L2-Sprecher mit B2- und C2-Sprachkompetenzniveau sowie sechs deutsche Muttersprachler Textpassagen aus dem EUROM-1-Korpus (Chan et al. 1995) auf Deutsch vorlasen.

3. Analyse des Surprisal-Effekts in sublexikalischen Einheiten

3.1. Auswirkungen des Surprisal-Effekts auf die Produktion und Wahrnehmung von Segmenten

3.1.1. Formant-Dynamik (Brandt et al. 2021)

Die Ausweitung des Vokalraums infolge von Vorhersagbarkeitseffekten lässt sich nicht nur in lokalen spektralen Messungen, sondern auch in den Formantverläufen erkennen. Vorhersagbare deutsche Vokale zeigen eine geringere Formantbewegung, eine flachere Steigung sowie eine langsamere Geschwindigkeit von F1 und eine geringere Krümmung in F2 im Vergleich zu weniger vorhersagbaren Vokalen. Vokale mit niedrigem Surprisalwert weisen geringere vokalspezifische spektrale Veränderungen auf, während Vokale mit hohem Surprisalwert mit einer größeren Dispersion im akustischen Vokalraum produziert werden (Brandt et al. 2019). Dies bedeutet, dass Sprecher Vokale in unvorhersagbaren Kontexten deut-

licher artikulieren, um ihre Unterscheidbarkeit zu erhöhen und die Übertragung wichtiger Informationen zu gewährleisten.

In dieser Studie wurde untersucht, inwieweit die Variabilität deutscher Formantverläufe durch Surprisal, Prominenz (primäre lexikalische Betonung) und deren Interaktion beeinflusst wird. Als zusätzliches informationstheoretisches Maß floss die Worthäufigkeit in die Modelle ein. Zur Analyse der Form von Formantverläufen in verschiedenen Surprisal-Kontexten kamen verallgemeinerte additive gemischte Modelle (GAMMs) zum Einsatz. Die Surprisalwerte basierten auf den Biphonen oder Triphonen des umgebenden Vokalkontexts, wobei nur Monophthonge in Inhaltswörtern berücksichtigt wurden.

Für die durchschnittlichen F1- und F2-Werte ergaben sich erwartungsgemäße Unterschiede zwischen den Vokalphonemen im akustischen Vokalraum. Ein signifikanter Interaktionseffekt zwischen Vokal und Betonung in den F1- und F2-Modellen zeigte, dass betonte Vokale eine größere Dispersion im Vokalraum aufweisen als unbetonte Vokale. Die Ergebnisse der GAMM-Analyse zeigen, dass die F1- und F2-Verläufe durch Surprisal, den umgebenden Kontext und die Interaktion von Surprisal und Betonung beeinflusst werden. Unbetonte Vokale zeigen eine größere Variabilität in ihren Formantverläufen bei verschiedenen Surprisalniveaus als betonte Vokale. Unterschiede in den durchschnittlichen Formantwerten lassen sich bei betonten Vokalen zudem leichter in Abhängigkeit vom Surprisalniveau modellieren als bei unbetonten Vokalen. Die Ergebnisse deuten darauf hin, dass die Effekte der kontextuellen Vorhersagbarkeit auf die Formantenvariabilität nicht nur auf lokale Messungen im Vokal begrenzt sind, wie es bei Studien zur Vokaldispersion der Fall war (z. B. Malisz et al. 2018), sondern auch die Dynamik über die gesamte Vokaldauer hinweg beeinflussen. Vor dem Hintergrund der Hypothese der einheitlichen Informationsdichte (uniform information density hypothesis, Levy / Jaeger 2006) erweitern unsere Ergebnisse das Konzept, dass rationale Sprecher über die gesamte Äußerung hinweg Optimierungsstrategien in der Sprachproduktion anwenden, um eine erfolgreiche Kommunikation zu gewährleisten. Dieses strategische Verhalten wirkt sich auch auf die Eigenschaften der Formantenverläufe aus und bleibt bestehen, selbst wenn für linguistische Faktoren kontrolliert wird, die bekanntermaßen die Formantenverläufe beeinflussen, wie z. B. Vokaldauer und phonologischer Kontext.

3.1.2. Stimmhaftigkeit der Plosive (Ibrahim et al. 2022a)

Silben in vorhersagbaren Kontexten haben oft eine kürzere Dauer als in unvorhersagbaren Kontexten, jedoch bleibt die Frage offen, ob Vorhersagbarkeit phonetische Hinweise auf phonologische Merkmale gleichmäßig beeinflusst. In dieser Studie wurde der Einfluss der silbenbasierten Vorhersagbarkeit auf die Dauerparameter des Stimmhaftigkeitskontrasts bei Plosiven im Deutschen unter-

sucht, insbesondere auf die Stimmlatenzzeit (Voice Onset Time, VOT) und die Verschlussdauer. Die Zielkonsonanten /b, p, d, k/ wurden in betonten CV-Silben in mehrsilbigen Wörtern innerhalb eines Satzkontexts analysiert. Die Silben traten entweder in Kontexten mit niedriger oder hoher Vorhersagbarkeit auf, basierend auf einem Trigramm-Silbenmodell. Die Analysen zeigten einen Interaktionseffekt der Vorhersagbarkeit und des Informationsstatus der Konsonanten auf die VOT, jedoch nicht auf die Verschlussdauer. Die selektiven Effekte auf VOT deuten darauf hin, dass Vorhersagbarkeit empfindlich auf die phonologische Relevanz eines sprachspezifischen phonetischen Merkmals reagiert.

3.2. Zusammenspiel von Lombard-Effekt und Surprisal in der Sprachproduktion (Ibrahim et al. 2022b)

Sprecher passen ihre Sprechweise in lauten Umgebungen an, indem sie klarer und deutlicher artikulieren, während sie in vorhersehbaren Kontexten eher dazu tendieren, sprachlichen Aufwand zu reduzieren und die Wortdauer zu verkürzen. Es ist unklar, wie diese beiden kommunikativen Anforderungen erfüllt werden. Obwohl es eine Vielzahl von Studien zu Strategien der Optimierung des Sprachsignals in lauten Umgebungen gibt (s. Cooke et al. 2014), haben sich diese Studien entweder auf den Lombard-Effekt oder auf den Vorhersagbarkeitseffekt konzentriert, während unsere Arbeit die bisherige Literatur erweitert, indem sie beide Faktoren zusammen untersucht.

Die Auswirkungen der Eigenschaften eines geräuschbelasteten Kanals auf die Sprachproduktion wurden untersucht, insbesondere die akustischen Realisierungen von Silben in vorhersagbaren im Vergleich zu unvorhersagbaren Kontexten bei unterschiedlichen Hintergrundgeräuschpegeln. Es wurden akustische Merkmale erfasst, die aus den Zielsilben extrahiert wurden, darunter Dauer, durchschnittliche Intensität, Intensitätsreichweite sowie die mediane Grundfrequenz. Zusätzlich wurden die Formanten F1, F2 und die Differenz dazwischen (F2–F1) für die Vokale innerhalb der Silben analysiert. Untersucht wurde, ob ein unzuverlässiger, geräuschbehafteter Kanal additive oder interaktive Effekte auf die Optimierung des Sprachsignals hervorruft, die durch Surprisal entstehen. Die Hypothese lautete, dass Lombard-Sprache und Surprisal interaktiv die akustischen Variablen beeinflussen würden, da Sprecher in geräuschvollen Umgebungen Einheiten mit hohen Surprisalwerten mehr verstärken als solche mit niedrigen Surprisalwerten. Diese Strategie würde es den Sprechern ermöglichen, sowohl informativ als auch effizient zu kommunizieren.

Wie aus der Literatur zur Lombard-Sprache (s. Junqua 1996 für eine Übersicht) zu erwarten war, war die Silbendauer länger, die durchschnittliche Intensität höher, die Intensitätsreichweite breiter und die Grundfrequenz (F0) insgesamt höher, wenn Hintergrundgeräusch vorhanden war. Der Einfluss des Geräuschpe-

gels war jedoch nur für die Intensität (sowohl durchschnittlich als auch im Umfang) sowie für die Grundfrequenz (F0) signifikant. Unerwarteterweise hatte der Geräuschpegel keinen Einfluss auf die Silbendauer. Der Effekt von Surprisal war signifikant nur für die Silbendauer und die Intensitätsreichweite, wobei Silben mit einem hohen Surprisalwert eine längere Dauer und eine größere Intensitätsreichweite aufwiesen (siehe Abb. 1).

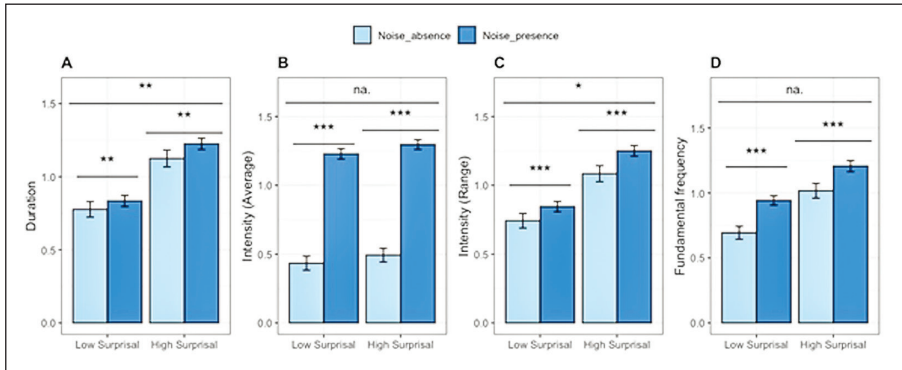


Abb. 1: Mittelwerte (Standardfehler) von silbenbasierten Merkmalen: (A) Dauer, (B) durchschnittliche Intensität, (C) Intensitätsbereich und (D) Grundfrequenz in Abhängigkeit von der Nichtvorhandensein (hellblau) bzw. Vorhandensein (dunkelblau) von Geräuschen (alle Werte z-standardisiert nach Sprecher).

Der Effekt von Geräusch war statistisch signifikant für die Formanten F1, F2 und ihre Differenz F2–F1 bei vorderen Vokalen. Vordere Vokale zeigten bei Vorhandensein von Geräusch höhere F1-Werte, niedrigere F2-Werte und kleinere F2–F1-Differenzen. Bei zentralen und hinteren Vokalen war der Effekt von Geräusch jedoch nur für F1 und F2–F1 signifikant, mit höheren F1-Werten und kleineren F2–F1-Differenzen bei Vorhandensein von Geräusch. Es wurden keine signifikanten Effekte von Surprisal oder Geräuschpegeln beobachtet. Diese Ergebnisse deuten darauf hin, dass Sprecher sich bemühen, den Unterschied zwischen Silben in Kontexten mit hohen bzw. niedrigen Surprisalwerten bei Vorhandensein von Hintergrundgeräusch zu verstärken. Darüber hinaus zeigen Lombard-Sprache und Vorhersagbarkeit additive Effekte auf die Verlängerung der Silbendauer und die Intensitätsreichweite, was darauf hindeutet, dass kanalspezifische Anpassungen vorgenommen werden, ohne die Kodierung der Informationsquelle zu beeinträchtigen. Es wurde keine Interaktion zwischen Geräusch und Vorhersagbarkeit festgestellt, was darauf schließen lässt, dass geräuschbedingte Modifikationen unabhängig von vorhersagbarkeitsbezogenen Veränderungen auftreten könnten.

Dies hat Implikationen für Sprachproduktionsmodelle, die sowohl kanalbasierte als auch inhaltsbasierte Modellelemente einbeziehen könnten. Obwohl die Kanal-kodierung nicht Teil der linguistischen Repräsentation während der Planungsphase ist, beeinflusst sie dennoch das phonetische Ergebnis. Unsere Studie untersuchte jedoch nur eine Art von Kanal. Es existieren zahlreiche andere Kanäle, beispielsweise die Kommunikation mit L2-Sprechern oder Robotern. Die Berücksichtigung der Kanaleigenschaften erfordert eine erweiterte Betrachtung der kommunikativen Kontexte als Teil der phonetischen Planung.

3.3. Effekte von Surprisal und prosodischer Struktur auf die Dauer der phrasenfinalen Silben (Andreeva et al. 2020)

Da die Prosodie bekanntermaßen informationstheoretische Effekte auf die phonetische Kodierung moduliert, wurde das mögliche Zusammenspiel zwischen Ebenen der prosodischen Hierarchie und dem Surprisal-Profil von Äußerungen untersucht. Damit sollte folgende Forschungsfrage beantwortet werden: Ist der Surprisal-Effekt vollständig der Prosodie zuzuschreiben, wie es die SSR-Hypothese in ihrer starken Interpretation nahelegt (Aylett / Turk 2004, 2006, Turk 2010), oder bleibt der Effekt zusätzlich zur Prosodie wirksam, wie unter anderem von Baker / Bradlow (2009) vorgeschlagen, oder besteht eine Interaktion zwischen Surprisal und Prosodie?

Untersucht wurden Effekte der prosodischen Struktur (Tonhöhenakzente und Grenzstärke) sowie von Surprisal auf die Dauer der finalen Silbe in einer Intonationsphrase (IP). Dabei wurden die Dauer der IP-finalen Silben und der darauffolgenden Pausen in einer Teilmenge des DIRNDL-Korpus (Eckart et al. 2012; Björklund, Börkelund et al. 2014) gemessen. Die Pausendauer diente als quantitativer Indikator für die Grenzstärke. Der Surprisalwert wurde anhand eines silbenbasierten Trigramms des vorangehenden Kontexts berechnet. Die Silbendauer wurde als Funktion von Surprisal und prosodischen Faktoren (Tonhöhenakzent und Grenzstärke) in linearen gemischten Modellen analysiert.

In Bezug auf die kontextuelle Vorhersagbarkeit zeigen unsere Ergebnisse, dass Surprisal die Dauer der IP-finalen Silbe vorhersagt. Die Analyse verdeutlicht, dass Silben, die schwer vorherzusagen sind, eine längere Dauer aufweisen im Vergleich zu leicht vorhersagbaren Silben. Diese Ergebnisse stehen im Einklang mit früheren Studien zur Dauer von Silben und ihrer Vorhersagbarkeit (Aylett and Turk 2004, Malizs et al. 2018, Gahl et al. 2012) und stellen eine wertvolle Erweiterung der Forschung zur Silbendauer im Rahmen der Informationstheorie dar. Besonders hervorzuheben ist, dass unsere Untersuchung sich auf IP-finale Silben konzentriert, welche bislang in der entsprechenden Forschungsliteratur weitgehend unbeachtet blieben. Des Weiteren konnten auch Akzentuierung und Grenzstärke als signifikante Faktoren für die Dauer der IP-finalen Silbe identifiziert werden. Wie erwartet, sind akzentuierte Silben länger als unakzentuierte,

und Silben, die vor starken Grenzen auftreten, haben eine kürzere Dauer im Vergleich zu denen vor schwachen Grenzen.

Es wurde weiterhin festgestellt, dass Akzentuierung und Surprisal als unabhängige Faktoren die Silbendauer beeinflussen. Interessanterweise zeigte Surprisal eine größere Effektstärke als die prosodischen Faktoren (die Akzentuierung und die Grenzstärke), was darauf hindeutet, dass die Modulation zwischen Surprisal und phonologischer Kodierung durch Prosodie nicht vollständig abgedeckt ist. Diese Ergebnisse unterstützen die schwache Version der SSR-Hypothese (Aylett / Turk 2004, 2006), da sie darauf hinweisen, dass die prosodische Struktur – insbesondere die Akzentuierung – nur wenige der signifikanten Effekte der Redundanz widerspiegelt. Mit anderen Worten, die Glättung scheint eher durch direkte Bezugnahme auf Redundanzfaktoren oder durch Grenzstärke zu erfolgen, was unter anderem eine Verlängerung der Dauer von phrasenfinalen Silben zur Folge hat.

Unsere Ergebnisse zeigen insgesamt eine Interaktion zwischen Surprisal und Grenzstärke, was darauf hindeutet, dass diese Faktoren sich bei der Erklärung der Variabilität in der Silbendauer ergänzen. Die Silbendauer nimmt mit steigendem Surprisal zu, wobei dieser Anstieg vor starken Grenzen signifikant größer ist als vor schwachen Grenzen (Abb. 2). Diese Modulation der Silbendauer tritt zusätzlich zum bekannten Effekt der phrasenfinalen Dehnung auf.

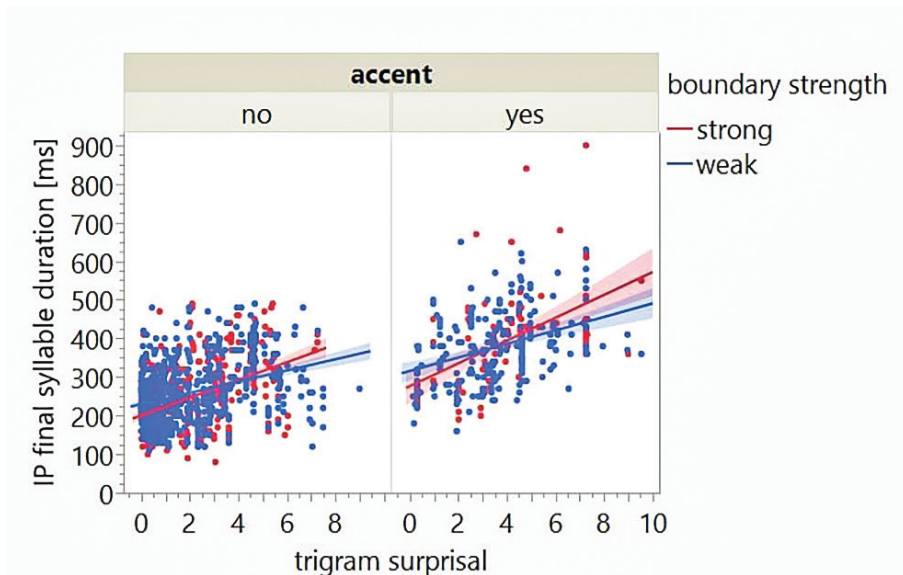


Abb. 2: Der Effekt von Trigramm-Surprisal, Grenztyp und Akzent auf die Dauer der finalen Silbe.

Die Ergebnisse dieser Studie heben die komplexen Interaktionen in der gesprochenen Sprache zwischen segmentalen, suprasegmentalen und informationstheoretischen Faktoren hervor, die mit der Vorhersagbarkeit im Kontext zusammenhängen.

3.4. Surprisal als Maß für Sprachkompetenz (Brand et al. 2019)

Die Untersuchung geht von der Annahme aus, dass Muttersprachler einer Sprache ein ähnliches Sprachmodell teilen, wobei eine gewisse individuelle Variabilität aufgrund von Idiolekt, soziolinguistischen oder regionalen Unterschieden besteht. Sprachlerner hingegen entwickeln durch ihre Auseinandersetzung mit der Zielsprache (L2) vermutlich mentale Modelle zur Vorhersagbarkeit sprachlicher Ereignisse in dieser Sprache. Diese Modelle variieren je nach Sprachkompetenz und Expositionsdauer des Lerners. Die Fähigkeit, bestimmte sprachliche Elemente vorherzusagen und sich an deren Variabilität anzupassen, steht im direkten Zusammenhang mit der Entwicklung von Sprachkompetenz. Die Studie untersucht, ob die Surprisal-Faktoren der Zielsprache (Deutsch) die phonetische Variabilität, insbesondere die Dispersion von Vokalen bei bulgarischen Deutschlernern unterschiedlicher Sprachkompetenzstufen, erklären können. Dies stellt einen innovativen, informationstheoretischen Ansatz dar, um den Spracherwerb aus der Perspektive der Vorhersagbarkeit zu untersuchen. Die zugrunde liegende Hypothese ist, dass Sprachlerner mit höherer Sprachkompetenz (C2) ein präziseres und stabileres mentales Modell der Zielsprache entwickeln, das in ihrer phonetischen Produktion erkennbar wird. Erwartet wurde, dass der Zusammenhang zwischen Surprisal und den Vokaldispersionsmustern, die bei L1-Sprechern beobachtet werden, auch bei fortgeschrittenen Lernenden (C2-Sprachkompetenzniveau) erkennbar ist, während er bei Lernenden mit mittlerem Sprachkompetenzniveau (B2) weniger stark ausgeprägt oder nicht vorhanden ist.

Die Vokaldispersion wurde für betonte gespannte und ungespannte Vokale in gelesenen Texten von deutschen Muttersprachlern und bulgarischen L2-Sprechern auf dem fortgeschrittenen (C2) und mittleren (B2) Sprachkompetenzniveau gemessen. Die Zusammenhänge zwischen der Vokaldispersion und den Surprisalwerten wurden für jede Sprechergruppe unter Verwendung der Pearson-Korrelation ausgewertet. Für die L1-Sprecher und die L2-Sprecher auf dem C2-Niveau ergaben sich signifikante Korrelationen zwischen Vokaldispersion und Surprisal, für die B2-Sprecher zeigte sich hingegen keine signifikante Korrelation. Dies könnte darauf zurückzuführen sein, dass die phonetischen Reduktionsstrategien bei diesen Lernenden noch nicht vollständig internalisiert wurden. Anschließend wurden für jede Sprechergruppe (L1-, C2- und B2-Sprecher) drei separate lineare gemischte Modelle (LMMs) berechnet, wobei Surprisal, Wortklasse,

durchschnittliche Vokaldauer und Gespanntheit der Vokale als feste Faktoren und Sprecher sowie Wort als zufällige Faktoren berücksichtigt wurden. Die Ergebnisse zeigten, dass bulgarische L2-Sprecher, insbesondere auf B2-Niveau, eine größere Vokaldispersion als deutsche L1-Sprecher aufwiesen (Abb. 3). Obwohl alle analysierten Vokale betont wurden, hoben die B2-Sprecher ihre Vokale bei niedrigen Surprisalwerten an, was eine Übertragung des bulgarischen Musters der unbetonten Vokalanhebung widerspiegelt (Andreeva et al. 2013, Sabev / Andreeva 2024). Dies lässt auf ein gewisses Bewusstsein der bulgarischen B2-Sprecher für die phonologischen Strukturen und Vorhersagbarkeit des Deutschen schließen, jedoch ohne die deutsche Reduktionsstruktur bei geringeren Surprisalwerten vollständig zu erreichen.

Zusammenfassend lässt sich sagen, dass das Ergebnis dieser Untersuchung nicht nur die Rolle von Surprisal als messbares Kriterium für die Sprachkompetenz unterstreicht, sondern auch die Bedeutung von Phonologie und prosodischen Strukturen in der sprachlichen Produktion von L2-Lernenden aufzeigt. Darüber hinaus zeigen die Ergebnisse der Studie auf, wie phonologische Modelle der L2 sich mit fortschreitendem Lernen und höherer Sprachkompetenz entwickeln, und wie diese Modelle mit den vorhersagbaren Mustern von sprachlichen Ereignissen interagieren.

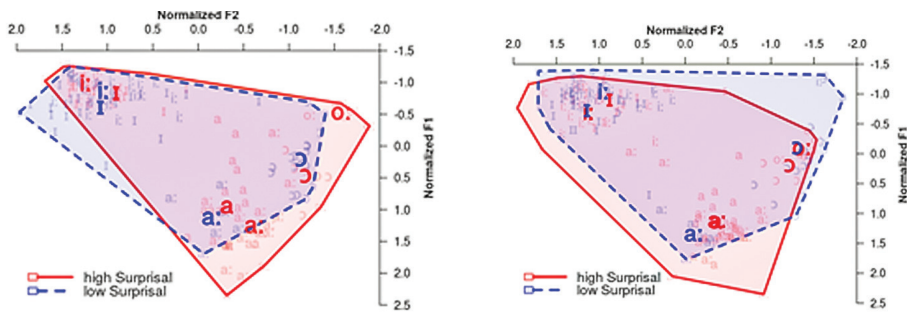


Abb. 3: Vokalsräume bulgarischer L2-Sprecher mit fortgeschrittenem (links) und mittlerem (rechts) Sprachkompetenzniveau. Vokale mit hohen Surprisalwerten sind in rot angegeben. Vokale mit niedrigen Surprisalwerten sind in blau angegeben.

4. Diskussion

Wenn phonetische Strukturen schwer aus ihrem Kontext vorherzusagen sind, erfahren sie eine Verstärkung im temporalen und spektralen Bereich. In gewissem Maße werden die Effekte der Vorhersagbarkeit durch die prosodische Struktur moduliert. Diese Forschung basiert auf der Annahme, dass Sprecher die Details des phonetischen Signals anpassen, um ein Gleichgewicht zwischen Sur-

prisal (struktureller Informationsdichte) und phonetischer Kodierung aufrechtzuerhalten. In dieser Arbeit wurden die Ergebnisse von Experimenten und Analysen von Korpora präsentiert, die darauf abzielten, diese Hypothese zu testen. Die Effekte von Surprisal auf die phonetische Kodierung wurden insbesondere in Bezug auf die dynamischen Formantverläufe von Vokalen (Brandt et al. 2021), die Stimmhaftigkeit von Plosiven (Ibrahim et al. 2022a), die Silbendauer und die Vokalraumausdehnung (Brand et al. 2019) untersucht. Dabei wurden verschiedene Faktoren, die mit der prosodischen Struktur zusammenhängen, wie lexikalische Betonung und prosodische Grenzen, als Kontrollfaktoren in den statistischen Modellen berücksichtigt, um die phonetische Wirkung von Veränderungen in Surprisal zu erfassen (Brandt et al. 2021). Zudem wurden mögliche Interaktionen zwischen Surprisal und der Optimierung des Sprachsignals unter geräuschvollen Bedingungen, bekannt als Lombard-Effekt (Ibrahim et al. 2022b), sowie zwischen Surprisal und der prosodischen Struktur (Andreeva et al. 2020) untersucht.

Die Untersuchung der Auswirkungen von Vorhersagbarkeit und Prominenz auf die dynamische Struktur der ersten und zweiten Formanten deutscher Vokale (Abschnitt 3.1.1, Brandt et al. 2021) zeigte, dass die Effekte kontextueller Vorhersagbarkeit auf feine phonetische Details nicht nur in lokalen Messungen, sondern auch in dynamischen Merkmalen phonetischer Segmente beobachtet werden können. Surprisal, basierend auf dem nachfolgenden Kontext, konnte die Formantenverlaufsgestalt im Deutschen erklären. Dieses Ergebnis war nicht unbedingt vorhersehbar, da die Richtung des Surprisal-Effekts stark von der untersuchten akustischen Dimension abhängt. Beispielsweise kann die Segmentdauer durch Surprisal des vorhergehenden und nachfolgenden Kontexts erklärt werden, während die Vokaldispersion nur durch Surprisal des vorhergehenden Kontexts vorhergesagt wird (vgl. Malisz et al. 2018).

Die Analyse des Effekts der silbenbasierten Vorhersagbarkeit auf die dauerbezogenen Korrelate des phonologischen Stimmhaftigkeitskontrasts im Deutschen (Abschnitt 3.1.2, Ibrahim et al. 2022a) deutete auf eine Interaktion zwischen Vorhersagbarkeit und Stimmhaftigkeit der Zielkonsonanten hinsichtlich der Stimmlatenzzeit hin, während ein einheitlicher Effekt auf die Verschlussdauer beobachtet wurde. Diese Erkenntnis unterstützt die Interpretation, dass der Effekt der Vorhersagbarkeit auf ein phonetisches Merkmal sensitiv für die phonologische Relevanz dieses Merkmals in der jeweiligen Sprache ist.

Die Ergebnisse der Studie zur Interaktion der Effekte von Surprisal und der Optimierung des Sprachsignals in geräuschvollen Umgebungen (Abschnitt 3.2, Ibrahim et al. 2022b) unterstützen die Interpretation, dass Sprecher sich bemühen, den Unterschied zwischen Silben in mehr oder weniger vorhersehbaren Kontexten bei Vorhandensein von Geräuschen zu vergrößern. Offensichtlich

haben Lombard-Sprache und Unvorhersagbarkeit additive Effekte auf die Erhöhung der Silbendauer und des Intensitätsbereichs, was darauf hindeutet, dass kanalbezogene Eigenschaften ausgeweitet werden, ohne die Quellenkodierung zu beeinträchtigen. Da keine Interaktion zwischen Geräusch und Surprisal gefunden wurde, könnten geräuschbezogene Modifikationen unabhängig von den durch Surprisal induzierten Effekten sein. Falls dem so ist, sollten Modelle der Sprachproduktion sowohl kanalbasierte als auch inhaltspezifische Formulierungen einbeziehen. Obwohl die Kanalkodierung nicht Teil der linguistischen Repräsentation während der Sprachplanung ist, prägt sie das phonetische Output.

Die Fallstudie zu Surprisal als Indikator für Sprachkompetenz (Abschnitt 3.4, Brand et al. 2019), die auf den Leseproduktionen fortgeschrittener und intermediärer bulgarischer Sprecher des Deutschen basiert, mit L1-Deutschsprechern als Kontrollgruppe, zeigt ein gewisses Maß an Bewusstsein bei intermediären L2-Sprechern für die deutschen phonologischen Strukturen und deren Vorhersagbarkeit. Dennoch gelang es diesen Sprechern nicht, das Reduktionsmuster der Zielsprache für Vokale in Kontexten mit niedrigen Surprisalwerten zu produzieren. Sie verließen sich stattdessen auf ihr L1-Reduktionsmuster. Im Gegensatz dazu wiesen die Vokalräume fortgeschrittener L2-Sprecher Dispersionseffekte auf, die denen von L1-Deutschsprechern ähnlicher waren. Dieses Ergebnis legt nahe, dass Sprachlerner mit zunehmender Exposition gegenüber der L2 mentale Modelle der Vorhersagbarkeit linguistischer Ereignisse in ihrer Zielsprache entwickeln.

Abschließend wurde der Effekt von Elementen der prosodischen Struktur, nämlich Tonhöhenakzenten und Grenzstärke, sowie von Surprisal auf die Dauer der letzten Silbe in der Intonationsphrase untersucht (Abschnitt 3.3, Andreeva et al. 2020). Die Ergebnisse zeigten eine Interaktion zwischen Surprisal und Grenzstärke in der letzten Silbe der Intonationsphrase. Mit steigenden Surprisalwerten nimmt die Dauer der Silbe zu, wobei dieser Effekt vor starken Grenzen ausgeprägter auftritt. Dies geht über den gut dokumentierten Effekt der phrasenfinalen Dehnung hinaus. Es wurde somit Evidenz für eine Interaktion zwischen Surprisal und Grenzstärke gefunden. Diese Faktoren ergänzen sich in der Erklärung der Variabilität der Silbendauer. Da Surprisal eine stärkere Wirkung als die prosodischen Faktoren hatte und die prosodische Struktur anscheinend nicht viel von den signifikanten Effekten der Redundanz umsetzen kann, könnte die Modulation zwischen Surprisal und phonetischer Kodierung durch Prosodie möglicherweise nicht umfassend sein. Daher sind die Ergebnisse allgemein mit einer schwachen Version der SSR-Hypothese (Aylett / Turk 2004, 2006) kompatibel.

Zusammenfassend deuten die Ergebnisse darauf hin, dass die prosodische Struktur zwischen den Anforderungen an effiziente Kommunikation und dem Sprachsignal vermittelt. Diese Vermittlung ist jedoch nicht perfekt. Es gibt Evi-

denz für zusätzliche, direkte Effekte von Veränderungen in der Vorhersagbarkeit auf die phonetische Struktur von Äußerungen. Darüber hinaus zeigen diese Effekte Stabilität bei variierenden Sprechgeschwindigkeiten in Modellen, die aus den Daten von sechs verschiedenen europäischen Sprachen abgeleitet wurden (Malisz et al. 2018).

Danksagung: Diese Forschung wurde von der Deutschen Forschungsgemeinschaft (DFG) gefördert, Projekt-ID 232722074 – SFB 1102 (Projekt C1).

LITERATURVERZEICHNIS/REFERENCES

- Andreeva, Bistra / Barry William J. / Koreman Jacques. 2013. “The Bulgarian stressed and unstressed vowel system. A corpus study.” In *Proceedings of Interspeech 2013* (345–348). <https://doi.org/10.21437/Interspeech.2013-97>
- Andreeva, Bistra / Möbius, Bernd / Whang, James. 2020. “Effects of surprisal and boundary strength on phrase-final lengthening. In *Proceedings of Speech Prosody 2020* (146–150). <http://dx.doi.org/10.21437/SpeechProsody.2020-30>
- Aylett, Matthew / Turk, Alice. 2004. “The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47: 31–56. <https://doi.org/10.1177/00238309040470010201>
- Aylett, Matthew / Turk, Alice. 2006. “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei.” *Journal of the Acoustical Society of America* 119 (1): 30–48. <https://doi.org/10.1121/1.2188331>
- Baker, Rachel A. / Bradlow, Anna R. 2009. “Variability in word duration as a function of probability, speech style, and prosody.” *Language and Speech* 52(4): 391–413. <https://doi.org/10.1177/0023830909336575>
- Baroni, Marco / Bernardini, Silvia / Ferraresi, Adriano / Zanchetta, Eros. 2009. “The WaCky Wide Web: a collection of very large linguistically processed web-crawled corpora.” *Language Resources and Evaluation* 43: 209–226. <https://doi.org/10.1007/s10579-009-9081-4>
- Björklund, Anders / Eckart, Kerstin / Riester, Arndt / Schauffler, Nadja / Schweitzer, Katrin. 2014. “The extended DIRNDL corpus as a resource for automatic coreference and bridging resolution.” In *Proceedings of LREC* (3222–3228). http://www.lrec-conf.org/proceedings/lrec2014/pdf/891_Paper.pdf [Zugriff am 28.02.2025]
- Brandt, Erika / Andreeva, Bistra / Möbius, Bernd. 2019. “Information density and vowel dispersion in the productions of Bulgarian L2 speakers of German.” In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)* (3165–3169). https://assta.org/proceedings/ICPhS2019/papers/ICPhS_3214.pdf [Zugriff am 28.02.2025]
- Brandt, Erika / Möbius, Bernd / Andreeva, Bistra. 2021. “Dynamic formant trajectories in German read speech: Impact of predictability and prominence.” *Frontiers in Communication / Language Sciences* 6: 1–15. <https://doi.org/10.3389/fcomm.2021.643528>
- Chan, Dominic / Fourcin, Adrian / Gibbon, Dafydd / et al. 1995. “EUROM- A Spoken Language Resource for the EU.” In *Eurospeech’95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology. Madrid, Spain, Vol 1* (867–870).

- Cooke, Martin / King, Simon / Garnier, Maëva / Aubanel, Vincent. 2014. "The listening talker: a review of human and algorithmic context-induced modifications of speech." *Computer Speech and Language* 28 (2): 543–571. <https://doi.org/10.1016/j.csl.2013.08.003>
- Eckart, Kerstin / Riestler, Arndt / Schweitzer, Katrin. 2012. "A discourse information radio news database for linguistic analysis." In *Linked data in linguistics: Representing and connecting language data and language metadata* (65–75), ed. by C. Chiarcos et al. Berlin/Heidelberg: Springer.
- Gahl, Susanne / Yao, Yao / Johnson, Keith. 2012. "Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech." *Journal of Memory and Language* 66 (4): 789–806.
- Hale, John. 2016. "Information-theoretical complexity metrics." *Language and Linguistics Compass* 10 (9): 397–412. <https://doi.org/10.1111/lnc3.12196>
- Ibrahim, Omnia / Yuen, Ivan / Andreeva, Bistra / Möbius, Bernd. 2022a. "The effect of predictability on German stop voicing is phonologically selective". In *Proceedings of Speech Prosody 2022* (669–673). <https://doi.org/10.21437/SpeechProsody.2022-136>
- Ibrahim, Omnia / Yuen, Ivan / van Os, Marjolein / Andreeva, Bistra / Möbius, Bernd. 2022b. "The combined effects of contextual predictability and noise on the acoustic realisation of German syllables." *Journal of the Acoustical Society of America* 152 (2): 911–920. <https://doi.org/10.1121/10.0013413>
- Junqua, Jean-Claude. 1996. "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the lombard reflex." *Speech Communication* 20 (1): 13–22.
- Levy, Roger / Jaeger, T. Florian. 2006. "Speakers optimize information density through syntactic reduction". In *Advances in neural information processing systems* 19 (849–856), ed. by B. Schölkopf et al. Cambridge: MIT Press. <https://doi.org/10.7551/mitpress/7503.003.0111>
- Malisz, Zofia / Brandt, Erika / Möbius, Bernd / Oh, Yoon Mi / Andreeva, Bistra. 2018. "Dimensions of segmental variability: interaction of prosody and surprisal in six languages." *Frontiers in Communication / Language Sciences* 3: 1–18. <https://doi.org/10.3389/fcomm.2018.00025>
- Mayer, Jörg. 1995. *Transcription of German intonation – the Stuttgart system*. Tech. rep. Institute of Natural Language Processing, University of Stuttgart.
- Möhler, Gregor / Schweitzer, Antje / Breitenbücher, Mark / Barbisch, Martin. 2000. *IMS German Festival* (version 1.2-os). University of Stuttgart: Institut für Maschinelle Sprachverarbeitung (IMS). [Zugriff am 02.01.2020].
- Sabev, Mitko / Bistra Andreeva. 2024. "The acoustics of Contemporary Standard Bulgarian vowels: A corpus study." *The Journal of the Acoustical Society of America*, 155 (3): 2128–2138.
- Shannon, Claude Elwood. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27: 379–423.
- Schiel, Florian. 1997. *Siemens Synthesis Corpus – SII000P*. University of Munich.
- Stolcke, Andreas. 2002. "SRILM — an extensible language modeling toolkit." In *Proceedings of Interspeech 2002* (901–904). https://www.isca-archive.org/icslp_2002/stolcke02_icslp.html [Zugriff am 28.02.2025].
- Turk, Alice. 2010. "Does prosodic constituency signal relative predictability? A Smooth

Signal Redundancy hypothesis.” *Laboratory Phonology* 1: 227–262. <https://doi.org/10.1515/LABPHON.2010.012>

Witten, Ian H. / Bell, Timothy C. 1991. “The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression.” *IEEE Transactions on Information Theory* 37 (4): 1085–1094.

✉ Prof. Dr. Bistra Andreeva
ORCID iD: 0000-0003-2774-1346
Language Science and Technology
Building C7.2, Room 5.02
66123 Saarbrücken, Germany
E-mail: andreeva@lst.uni-saarland.de

✉ Prof. Dr. Bernd Möbius
ORCID iD: 0000-0003-3065-9984
Language Science and Technology
Building C7.2, Room 4.10
66123 Saarbrücken, Germany
E-mail: moebius@lst.uni-saarland.de

✉ Dr. Ivan Yuen
ORCID iD: 0000-0002-3238-0402
Language Science and Technology
Building C7.2, Room 5.03
66123 Saarbrücken, Germany
E-mail: ivyuen@lst.uni-saarland.de

✉ Dr. Omnia Ibrahim
ORCID iD: 0000-0002-3649-7376
Language Science and Technology
Building C7.2, Room 5.06
66123 Saarbrücken, Germany
E-mail: omnia@lst.uni-saarland.de