

Vorgehen bei der Analyse für die Artikelschreibung (Wörterbuch)



Gabriele Langer

DOI (jeweils letzte Version): [10.25592/uhhfdm.9576](https://doi.org/10.25592/uhhfdm.9576)

Version 1	2016-12-20	Erste Fassung
Version 2	2017-04-03	Version vom Zwischenbericht 2016
Version 3	2020-07-17	Kleinere Korrekturen im Text zum mittlerweile leicht geänderten Vorgehen sind eingearbeitet (Stand: Sommer 2020). Insbesondere wurden die Sichten im Anhang, die sich weiterentwickelt haben, und die Verweise auf relevante Veröffentlichungen unsererseits auf den aktuellen Stand gebracht. Die Screenshots der Beispiele sind aber teilweise noch Stand 2016/2017 (aus Version 1), da die Datenlage z.T. überholt ist und sich die Beispiele in iLex aktuell ggf. anders darstellen dürften. Die prinzipiellen Punkte, die die Beispiele illustrieren, werden jedoch auch durch die alten Screenshots deutlich. Ergänzt: Englisch Abstract, Links bei Referenzen.

Zusammenfassung

Seit 2016 werden laufend Wörterbucheinträge für das korpusbasierte Wörterbuch erarbeitet. Hier wird der Workflow zur Analyse und Bearbeitung der Wörterbucheinträge in einzelnen Arbeitsschritten aufgelistet und kurz beschrieben. Im Anhang finden sich Beispiele der Bearbeitungsschritte und verschiedene Sichten auf die Daten, die zur Bearbeitung genutzt werden.

Abstract

Beginning in the Spring of 2016 dictionary entries for the corpus-based dictionary are being compiled. For this a workflow has been developed. In these Project Notes the different steps of the workflow for analyzing and compiling dictionary entries for the corpus-based Dictionary are listed and shortly described. Main steps are 1. selection of lemma sign candidates, 2. review of type structures and type surroundings in the annotational database iLex, 3. establishment of lemma signs, 4. cross-references on entry level, 5. Review of collocational and phrase/compound patterns, 6. word sense discrimination (WSD), 7. resolving open questions, 8. selection and summary of senses for production, 9. selection of examples from the corpus and preparing them for publication, 10. final completion and editing of entry, 11. production of videos, 12. Production of entries for online display. The appendix contains examples of workflow steps and different views on the data that are used in the making.

Inhaltsverzeichnis

Hintergrund	3
Vorausgehende Arbeiten	4
Sichten auf die Daten	4
Analyse	5
1. Auswahl (Lemmaselektion)	5
2. Sichtung: Überblick über Typestruktur und Typeumgebung	7
3. Lemmazeichen-Konstituierung.....	8
4. Verweise auf Eintragebene	8
5. Überblick über Kollokationen und Zusammensetzungen	9
6. Lesarten-Differenzierung (Word Sense Discrimination WSD)	9
6a. Überblick über die Beleglage und Auswahl der zu bearbeitenden Tokens	10
6b. Sichten und Analyse einzelner Belege.....	10
6c. Gruppierung der Belege und Anlage der Proto-Lesarten.....	11
6d. Einstufung der Lesarten.....	11
7. Klärung offener Fragen	11
7a. Detailtranskription beauftragen.....	11
7b. Fokusgruppe.....	12
7c. DGS-Feedback.....	12
7d. Nacherhebungen.....	12
7e. Redaktionssitzung	12
Artikelschreibung	12
8. Auswahl und Zusammenfassung der Lesarten fürs Produkt	12
9. Korpusbeispiele für die Produktion auswählen und bearbeiten	13
10. Alle Angaben vervollständigen und in Form bringen	13
Produktion	14
11. Videoproduktion	14
11a. Regielistenerstellung	14
11b. Studioaufnahmen.....	14
11b. Postproduktion der Studioaufnahmen	15
12. Produktion der Vorabenträge	15
References	15
Anhang	17

Hintergrund

Die Inhalte des Wörterbuchs (im Folgenden auch als WB abgekürzt) sollen so weit als möglich korpusbasiert sein. Alle Angaben zur Verwendung einer Gebärde werden auf der Grundlage der zur Verfügung stehenden transkribierten Korpusdaten erarbeitet. Das bedeutet im Umkehrschluss, dass zunächst insbesondere Verwendungen (Bedeutungen/Lesarten, aber auch Formen), die nicht im Korpus enthalten sind, nicht Eingang in die vorläufigen Wörterbucheinträge finden, auch wenn sie existieren und vielleicht sogar geläufig und verbreitet sind. In dieser Hinsicht ist das WB von der Anlage her sogar weitgehend korpusgebunden.

In besonderen und klar definierten Fällen werden die korpusbasierten Angaben darüber hinaus jedoch noch durch andere Quellen oder Daten ergänzt.¹ Einträge und Angaben in Einträgen, die aus anderen Quellen stammen und nicht oder nur sehr eingeschränkt auf Korpusdaten beruhen oder durch Korpusdaten abgesichert sind, werden im Wörterbuch als solche gekennzeichnet, damit für die Nutzer_innen jederzeit erkennbar ist, welche Angaben über Korpusdaten abgesichert sind und welche nicht.

Bisher sind wir in folgenden Fällen über das Korpus hinausgegangen:

- Die im Zuge der Bearbeitung des Grundwortschatzes bereits durch das Feedback gewonnenen Daten und Erkenntnisse zu einzelnen Gebärden und deren Formen und Bedeutungen werden bei der Analyse mitberücksichtigt und runden das Bild, das sich in den Korpusdaten zeigt, ab. Ergebnisse aus dem DGS-Feedback werden zum Beispiel zur Einschätzung einzelner, z.T. im Korpus nur schwach belegter Lesarten herangezogen. Im Korpus sehr schwach belegte, im Feedback aber klar verbreitete Kernlesarten werden im WB entsprechend berücksichtigt. Die Ergebnisse des DGS-Feedbacks werden darüber hinaus auch genutzt, um Angaben zur regionalen Verbreitung klarer analysieren zu können, als dies mit 330 Informant_innen möglich ist (s. Wahl et al. 2018).
- Wenn es konkrete Einzelfragen zur Verwendung vor allem regionaler Gebärden gibt, wird gelegentlich auch auf das Wissen der Fokusgruppe bzw. des jeweiligen regionalen Vertreters zurückgegriffen. Die Fokusgruppenmitglieder helfen auch durch ihr Sprachwissen, synonyme und antonyme Gebärden in den Einträgen zu überprüfen und zu ergänzen.
- Regional stark begrenzte Gebärden weisen i.d.R. kleinere Tokenzahlen auf, hier weichen wir ggf. von dem Bearbeitungsschwellenwert von ≥ 25 ab und bearbeiten Einträge auch auf Grundlage von kleineren Tokenzahlen, wenn diese aussagekräftig sind. Dies gilt besonders bei Gebärden für begrenzte Sets einer bestimmten semantischen Kategorie (siehe nächster Punkt).
- Bei begrenzten Sets von Gebärden für eine bestimmte semantische Kategorie, wie z.B. Monate, Wochentage und Farben, die darüber hinaus auch eine starke regionale Auffächerung erfahren, wird zum einen die Abdeckung des gesamten Sets angestrebt, wie auch die Berücksichtigung von Gebärden aus möglichst allen Regionen. Dabei werden ggf. auch schwach und sehr schwach belegte Types – ggf. unter Hinzunahme anderer Quellen wie dem DGS-Feedback – berücksichtigt.
- Auch bei deutschen Städten haben wir uns aus Gründen der Abdeckung entschieden, über die im Korpus vorhandenen Gebärden hinauszugehen. Hier haben wir die Möglichkeit gezielter Nacherhebungen genutzt. Dafür wurde das Programm SignHunter entworfen und eingesetzt (s. Hanke et al. 2020).

¹ Im Antrag sind dazu die Task *T10.8 Klärung von Detailfragen (Nacherhebung)* mit den Möglichkeiten introspektiver Diskussionsrunden mit den gehörlosen Mitarbeitern und Mitarbeiterinnen und der Fokusgruppe sowie kleinerer Nacherhebungen und die Tasks *T10.9 Daten für das Voting erstellen und aufbereiten* und *T10.10 Auswertung des Votings* vorgesehen (s. Projektantrag S. 75/76).

Im Folgenden werden die Arbeitsschritte und der Workflow² skizziert, wie er sich für den korpusbasierten Anteil der Analyse darstellt, ohne die Abrundung durch weitere Informationsquellen wie dem DGS-Feedback oder Nacherhebungen mit einzubeziehen.³

In der jetzigen Projektphase wird die Bearbeitung der Einträge so weit vorangetrieben, dass sie inhaltlich einen brauchbaren Reifegrad erreichen und als Vorabeinträge schon eine sinnvolle punktuelle Informationsquelle zu den bearbeiteten Gebärden darstellen. Gleichzeitig lassen die Einträge erkennen, welche Art von Angaben das Wörterbuch bieten wird. Die Vorabeinträge erheben jedoch nicht den Anspruch, schon vollständige Einträge hinsichtlich der Art der Informationseinheiten, dem Umfang der enthaltenen Lesarten und der Art der Präsentation der Informationen (Layout, Zugriffswege, endgültige Formulierungen etc.) zu sein.

Vorausgehende Arbeiten

Der Workflow sieht vor, dass die für das WB zu bearbeitenden Gebärden und deren Varianten vorher den Arbeitsschritt der Lemmarevision („Glossenprüfen“) durchlaufen haben. Dieser Arbeitsschritt soll sicherstellen, dass die Lemmatisierung (Token-Type-Zuordnung) und die Differenzierung in Gebärdenformen („Wortformen“ und „Varianten“) zuverlässig und belastbar sind. In der Praxis gelingt dies nicht immer. Manchmal ergeben sich bei der lexikographischen Analyse auch neue Fragestellungen an die Transkription, die dann bei dieser zur Bearbeitung bestellt werden („Feintranskription“) und die auch die Lemmatisierung betreffen können.

Sichten auf die Daten

Voraussetzung für die Sichtung und Analyse der Daten bei der Artikelschreibung sind nützliche Sichten auf die verschiedenen Daten (Korpus/Feedback).

2016 wurden die ersten Lemmazeichen-Kandidaten allein auf Grundlage der Korpusdaten bearbeitet. Dabei wurden Abfragen und Sichten auf die Daten in iLex erstellt oder angepasst, die für diesen Arbeitsschritt benötigt oder sinnvoll sind. Es handelt sich um Darstellungen in iLex, die per SQL-Abfragen direkt aus dem aktuellen Datenbestand erzeugt werden. Sie werden je nach Bedarf überarbeitet und erweitert und können auch von der Bearbeiterin nach eigenen Wünschen selbst angepasst werden.⁴

Die als vorläufiges Dictionary Writing System (DWS) dienende FileMaker-Datenbank (im Folgenden FMP) wurde entsprechend dem korpusbasierten Verfahren erweitert und so weit wie möglich mit Scripts versehen, um die für die Analysearbeit benötigten Angaben per Abfrage direkt aus iLex abrufen zu können (per SQL-Abfrage).⁵

² Mit der Entwicklung der Analysemethoden und des Workflows für die rein korpusbasierte Analyse und Artikelschreibung wurde bereits vor 2016 zumindest in experimenteller Form begonnen. Auch schon zu Grundwortschatzzeiten wurde mehr und mehr auf Korpusdaten zurückgegriffen und für einzelne Type-Einträge der Annotationsdatenbank iLex eine rein korpusbasierte Analyse versucht, ohne dass dafür jedoch bereits die nötigen Strukturen zur Verfügung gestanden hatten wie z.B. entsprechende Strukturen in der FileMaker-Datenbank.

³ Zur Einbeziehung der Daten aus dem Feedback siehe Wähl et al. (2018).

⁴ Siehe auch Langer et al. (2018a).

⁵ FMP erfüllt größtenteils die Aufgaben des noch nicht vorhandenen Dictionary Writing Systems (DWS), das als Erweiterung von iLex geplant ist. FMP unterstützt einige Arbeitsschritte mittlerweile ganz gut, bleibt aber durch die nicht unmittelbare Anbindung an iLex und die dadurch entstehenden Datendiskrepanzen und notwendigen regelmäßigen Abgleichungsschritte bei der täglichen Arbeit umständlich. Einige Funktionen des DWS erfüllt FMP gar nicht, da sie ursprünglich nur als kurz- bis mittelfristige Notlösung gedacht war und weitergehende Funktionen bei der Einrichtung der Datenbank entsprechend nicht berücksichtigt wurden. Allerdings wird die Datenbank

Beispiele für verschiedene Sichten auf die Daten in iLex und die Arbeitsumgebung in FMP sind im Anhang zu diesem Arbeitspapier zusammengestellt.

Analyse

Im Folgenden werden das bisher erarbeitete Vorgehen für die Analyse dargestellt und die einzelnen, zu erledigenden Arbeitsschritte nacheinander beschrieben oder zumindest kurz angedeutet. Einige dieser Schritte erfolgen bei der Analyse jedoch mehr oder weniger parallel.

1. Auswahl (Lemmaselektion)

Bei der Lemmaselektion wird festgelegt, welche Gebärden Eingang in das Wörterbuch finden und in Einträgen beschrieben werden. Im Wesentlichen erfolgt das für das WB auf Grundlage der Korpusdaten (frequenzbasiert), ggf. mit systematisch ausgleichenden Ergänzungen. Im ersten Schritt werden Simplex-Gebärden beschrieben, die sich mit ausreichender Beleglage in den DGS-Korpus-Daten finden.

Zunächst nicht berücksichtigt werden einige sprachlichen Einheiten, die möglicherweise eine besondere Form der Darstellung benötigen und für die erst später entschieden werden soll, wie diese Darstellung am besten aussieht, d.h. ob sie in Form eigener Einträge, übersichtlich in einem Anhang oder in der Wörterbuchgrammatik beschrieben werden sollen:

- semantische Klassifikator-Handformen;
- andere produktiv einsetzbare Handformen⁶;
- Umgang mit Mundgestik und Mimik als „Building Blocks“⁷ analog zu Klassifikatoren: Welche dieser Einheiten werden am besten in der Wörterbuchgrammatik beschrieben? Welche sollten als sprachliche Einheiten oder Elemente sprachlicher Einheiten einen Eintrag im WB bekommen?
- Umgang mit Gebärden, die im Wesentlichen oder ausschließlich aus einer Mundgestik bestehen (wie z.B. LÜGEN);
- Mehrwortausdrücke⁸ (MWE): Zurzeit werden nur Gebärden-Simplexe bearbeitet. Wichtige MWE, die ein Element haben, das zu einer Gebärde der Kandidatenliste gehört, werden als zum Eintrag gehörig notiert. Eher transparente und parallel zum Deutschen gebildete Zusammensetzungen, z.B. als (Lehn-)Komposita eingeschätzte Kombinationen, werden in Auswahl als Run-ons am Ende des Eintrags aufgelistet. Eher DGS-spezifische „idiomatische“ Phrasen, die mehr Erklärungen oder Angaben erfordern, werden als Phrasen innerhalb, aber am Schluss der Lesartenliste beschrieben.

Die Reihenfolge der Bearbeitung und die Aufnahme als Eintrag in das Wörterbuch erfolgt im Wesentlichen frequenzbasiert, d.h. dass zunächst die Gebärden als Lemmazeichen-Kandidaten ausgewählt und bearbeitet werden, die im Korpus (zum aktuellen Bearbeitungsstand) so gut belegt sind, dass mit einer Lesartendifferenzierung (Word Sense Discrimination, WSD) aufgrund der Datenlage sinnvoll begonnen werden kann.

kontinuierlich weiterentwickelt und je nach Entwicklungsstand der WB-Inhalte und -Strukturen mit neuen Feldern und Funktionen versehen.

⁶ Vergleiche Brien (1992), Johnston (1989), Centre for Sign Language (2008-2017).

⁷ Vergleiche z.B. Johnston/Schembri (1999, 118).

⁸ Die für Gebärdensprachen passendere Bezeichnung Mehrgebärdenausdruck hat sich in der Gebärdensprachforschung bzw. -lexikographie nicht etabliert.

Als vorläufige Bearbeitungskandidaten betrachten wir Gebärden⁹ (iLex: types.level=3), die Lexeme (iLex: types.level=1) enthalten, die mit mindestens 25 Tokens belegt sind. Dieser Grenzwert dient als grobe Richtschnur und ist relativ willkürlich gewählt. Laut Sinclair (2004) braucht ein Lexikograph mindestens 20 Tokens eines nicht mehrdeutigen Wortes, um eine Vorstellung zu bekommen, wie dieses Wort verwendet wird.¹⁰ Wenn wir davon ausgehen, dass die Vorsortierung aus der Transkription, die mit der Lexem-Ebene dokumentiert ist, bereits eine grobe Trennung nach möglichen unterschiedlichen Bedeutungen (die sich in der Kombination der Gebärde mit verschiedenen Mundbildern widerspiegelt) darstellt, dann sollte ein Lexem mit einer Beleglage von ≥ 25 Tokens eine erste sinnvolle Analyse erlauben. Damit ist natürlich nicht automatisch sichergestellt, dass alle wichtigen Lesarten abgedeckt sind. Insofern stellen eine solche Analyse und damit auch die korpusbasierten Wörterbucheinträge, die daraus entstehen, immer ein vorläufiges Ergebnis dar, das ggf. später bei einer besseren Datenlage ergänzt werden kann. Wenn wir aber davon ausgehen, dass unser Korpus wesentlich kleiner ist als entsprechende Lautsprachkorpora und wir daher für viele Gebärden nur eher kleinere Belegmengen zur Verfügung haben, erscheint ein erster Richtwert von 25 Tokens pro Lexem brauchbar.

Aus dieser Vorüberlegung heraus ergibt sich die vorläufige Lemmazeichen-Kandidatenliste.¹¹ Innerhalb dieser Kandidatenliste gibt es keine feste Bearbeitungsreihenfolge. Sonderfälle wie \$INDEX können zunächst übersprungen werden. Die Bearbeiterinnen suchen sich geeignete

⁹ Die Type-Struktur in iLex ist als Type-Hierarchie angelegt. Jeder Type-Eintrag gehört zu einer Ebene (types.level). Auf der obersten Ebene (level=3) sind die „Gebärden“, die die Lemmazeichen-Kandidaten darstellen. Von den Gebärden abhängige Types sind den Ebenen 2, 1 und 0 zugeordnet und werden im Folgenden Subtypes genannt. Subtypes der Ebene 1 (types.level=1) werden in iLex als „Lexeme“ bezeichnet. Lexem-Types dienen dazu, Tokens konventionalisierter Form-Bedeutungs-Beziehungen, i.d.R. übliche Gebärde-Mundbild-Kombinationen, innerhalb eines Types zusammen zu gruppieren. Die Bezeichnung „Lexeme“ hier ist eine projektinterne und versteht sich nicht als lexikographischer oder lexikologischer Fachbegriff, der je nach Auffassung unterschiedlich besetzt sein kann. Vergleiche dazu Schlaefter (2002, 16): „Das Verständnis des Lexems als Systemwort mit Ausdrucks- und Inhaltskomponente wird lexikologisch in unterschiedlicher Weise spezifiziert. Während einerseits eine bestimmte Bedeutung als zeichenbildend betrachtet wird und damit jedes Lexem nur eine Bedeutung besitzt, gehen andere Auffassungen davon aus, daß Lexeme mehrere Bedeutungen haben können.“ (Vgl. in diesem Zusammenhang auch die Verwendung von „lexeme“ in Svensén (2009, z.B. 95f.). Atkins/Rundell (2008, 162) sprechen dagegen von „lexical units“ und „headwords“: „A headword in one of its senses is a lexical unit (or LU).“

Die Bezeichnung Lexem wurde eher der ersten Auffassung entsprechend vergeben, stellt in iLex aber lediglich eine Art Vorgruppierung dar. Im Rahmen der Token-Type-Zuordnung werden Tokens verschiedenen Subtypes zugeordnet. Die einzelnen Type-Ebenen erfüllen unterschiedliche Gruppierungsfunktionen: Die Subtypes der Ebene 1 (Lexeme) gruppieren Tokens nach ihren lexikalisierten Verwendungen (Bedeutungen), während die Subtypes der Ebene 0 bestimmte Realisierungsformen dieser lexikalisierten Verwendungen erfassen, um z.B. Wortformen zu gruppieren. Entsprechend gibt es auch weitere Realisierungsformen des Types, die keiner lexikalisierten Verwendung entsprechen und als Subtype der Ebene 2 angelegt werden. Jedes einem Subtype zugeordnete Token ist ein Beleg für diesen Subtype, aber auch gleichzeitig für alle ihm übergeordneten Subtypes bzw. Types.

¹⁰ Man kann nicht genau sagen, wie viele Belege man jeweils benötigt, um ein Wort lexikographisch gut beschreiben zu können (vgl. Atkins & Rundell 2008, S. 61). Sinclair spricht von einem Minimum von 20 Belegen: „Let us suggest some outline figures that may guide practice. A word which is not specially ambiguous will require at least twenty instances for even an outline description of its behaviour to be compiled by trained lexicographers.“ (Sinclair 2004, Abschnitt: 7. Size)

Aus statistischen Gründen ist für Lexikographen alles unter 5 Belegen nicht beachtenswert (vgl. Heid 2008, 135). Atkins & Rundell diskutieren, dass für ein Wort wie *adjudicate* 121 Belege ausreichend, 20 jedoch zu wenig wären, um die Bandbreite des grammatischen Verhaltens von *adjudicate* zuverlässig zu erkennen (vgl. Atkins & Rundell 2008, 60). Software, die aus mit Wortarten getaggen Korpusdaten (POS-Tagging) lexikalische Wortprofile erstellen kann (lexical profiling software) wie z.B. *Wordsketch*, benötigt laut Atkins & Rundell mindestens 500 Korpusbelege, um zuverlässige Ergebnisse zu liefern (vgl. Atkins & Rundell 2008, 61).

¹¹ Kandidatenliste Stand: 20.10.2016: 1139 Gebärden, denen 1613 Lexeme mit einer Beleglage von ≥ 25 Tokens zugeordnet sind; Stand 1.3.2020: 1538 Gebärden mit 5208 Lexemen, davon 2413 \geq mit 25 Tokens.

Gebärden zur Bearbeitung aus der Kandidatenliste aus. Was zuerst bearbeitet wird, ist einerseits abhängig von einer hohen Tokenzahl (gute Beleglage), ergibt sich aber auch aus dem Bearbeitungsstand der Lemmarevision und der Art der Gebärde.¹² Darüber hinaus werden im Korpus enthaltene Ausführungsvarianten gemeinsam bearbeitet (auch wenn nicht alle Varianten eine Beleglage von ≥ 25 aufweisen). Weiterhin kann es sinnvoll sein, (teil)synonyme oder bedeutungsähnliche Gebärden zeitlich aufeinanderfolgend zu bearbeiten, da sie ähnliche Lesartenstrukturen aufweisen können. Mit diesem Vorgehen wird auch gefördert, dass Querverweise auf bedeutungsähnliche Gebärden (Synonyme) in schon bearbeiteten Einträgen schneller zu ebenfalls bearbeiteten Zieleinträgen führen, anstatt auf lediglich automatisch generierte Einträge zu verweisen. Darüber hinaus werden semantisch zusammenhängende Sets auch bevorzugt zusammen bearbeitet.¹³

Insbesondere bei semantischen Sets findet man eine hohe regionale Variation mit regional stark begrenzten Varianten und entsprechend niedrigeren Tokenzahlen. Um eine gute regionale Abdeckung zu erreichen, wird bei Kandidaten für Varianten, die nur in einer begrenzten Region verwendet werden, von der Mindestzahl von 25 Tokens ggf. abgewichen (kleinerer Informantenkreis), insbesondere wenn sich mehrere regionale Varianten für verschiedene Regionen abzeichnen.

In der praktischen Arbeit wird so vorgegangen: Die Bearbeiterin wählt eine Gebärde für die Bearbeitung aus der Kandidatenliste aus. Im Idealfall ist für diese Gebärde und ihre Type-Umgebung die Lemmarevision bereits abgeschlossen. Ist dies nicht der Fall, kann zunächst die Lemmarevision in Auftrag gegeben werden, bevor mit der Arbeit begonnen wird.

Ein zur Bearbeitung ausgewählter Type wird in iLex mit einem entsprechenden Metadatum mit der Bearbeiterin versehen und in der Filemaker-Datenbank wird ein Eintrag angelegt, in dem der Lemmakandidat per drag and drop aus den zugehörigen iLex-Type eingetragen wird.

2. Sichtung: Überblick über Typestruktur und Typeumgebung

Im ersten Schritt verschafft sich die Bearbeiterin einen Überblick über die Gebärde, deren Umgebung und die Beleglage, um festzustellen, welche Types und Subtypes bei der Bearbeitung berücksichtigt werden müssen.

Zur Sichtung gehört:

- Überblick über die enthaltenen Lexeme (und eingetragenen konventionellen Bedeutungen), Überblick über Beleglage im Korpus: erster Eindruck zum Bedeutungsumfang; gibt es mehrere Varianten auf Lexem-Ebene in einem Type?
- Überblick über die nachgewiesenen Gebärdenformen (Qualified Types, Varianten),
- Überblick über Varianten (durch Glossierungskonvention als Varianten gekennzeichnete andere Gebärden bzw. Lexeme) und enthaltene Lexeme von Variantengebärden (s. Anhang Folien 3-7),
- Formgleiche und formähnliche Gebärden: Handelt es sich um verwandte Gebärden (\Rightarrow Verweis) oder Varianten (\Rightarrow im Artikel berücksichtigen) oder unabhängige Gebärden (\Rightarrow ggf. Verweis) oder gar „Wortformen“ (\Rightarrow berücksichtigen)? (s. Anhang Folie 9),

¹² Es ist sinnvoll, frühzeitig Erfahrungen mit verschiedenartigen Gebärden zu sammeln, die unterschiedliche analytische und lexikographische Bearbeitungsschritte erfordern, um entsprechende Erfahrungen in die weitere konzeptionelle Arbeit am WB einbringen zu können. Sehr hochfrequente Gebärden können zunächst übersprungen werden, um sich bei ihrer Analyse bereits Bearbeitungserfahrungen und NLP-Verfahren, die nach und nach dafür entwickelt werden, zu Nutze machen zu können (siehe z.B. Langer & Schulder 2020).

¹³ Es bietet sich ebenfalls an, Gebärden mit bestimmten grammatischen Eigenschaften, wie z.B. die Gruppe der so genannten Richtungsgebärden, gemeinsam zu bearbeiten, wenn bereits Erfahrungen mit einzelnen Vertretern dieser Gruppe gesammelt wurden.

- Gebärden und Lexeme mit gleichem Glossennamen (Types bzw. Subtypes-Liste): Handelt es sich um Ausführungsvarianten oder Gebärdenformen? (⇒ berücksichtigen) oder um synonyme/teilsynonyme Gebärden? (⇒ ggf. Verweise) oder um verwandte Types (⇒ ggf. Verweise)? (s. Anhang Folie 3),
- Für alle enthaltenen Lexeme: Überblick über Variantensets verschaffen (gleiches Set ⇒ in einem Eintrag darzustellen, verschiedene Sets ⇒ in getrennten Einträgen zu beschreiben),
- ggf. Überblick über regionale Verteilung (bei Verdacht auf Regionalität, s. Anhang Folie 4),
- ggf. Überblick über Feedback-Rückläufe (falls bereits im Feedback).

Ergeben sich aus dieser Sichtung noch Fragen an die Transkription, werden die betroffenen Gebärden oder Lexeme zur nochmaligen Bearbeitung zurück an die Lemmarevision verwiesen. Besteht der Bedarf, die Tokens noch nach zusätzlichen Gesichtspunkten zu taggen, kann beim Transkriptionsteam ein gezielter Annotationsauftrag für eine weitergehende Detailtranskription angefordert werden. Die Sichtung ist die Voraussetzung und damit auch schon Teil der Lemmakonstituierung.

3. Lemmazeichen-Konstituierung

Auf Grundlage der Sichtung und der Lemmatisierungsregeln¹⁴ für das WB wird festgelegt, welche Types und Subtypes zusammen in einem Eintrag beschrieben werden und welche Types und Subtypes des Lemmazeichen-Kandidaten und der Type-Umgebung in eigenen Einträgen beschrieben werden sollen.¹⁵ Damit wird insgesamt festgelegt, welchen Ausschnitt der Korpusdaten dieser Eintrag abdecken und beschreiben soll.

Für den so definierten Eintrag wird nach Sichtung der Datenlage festgelegt:

- welche Ausführungsvarianten aufgeführt werden,
- welche Variante als Hauptvariante gezeigt wird,
- welche Wortform der Hauptvariante als Grundform angesehen wird und damit als Zitatform (Lemma) das Lemmazeichen repräsentieren soll.

Das Ergebnis der Entscheidungen bei der Lemmakonstituierung sind angelegte Proto-Einträge¹⁶ im FMP, die in den folgenden Arbeitsschritten weiter befüllt werden.

4. Verweise auf Eintragungsebene

Auf Grundlage der Sichtung der Type-Umgebung und der Entscheidungen zur Lemmazeichen-Konstituierung werden folgende Informationen im Eintrag festgehalten:

- Verweise auf formgleiche, formähnliche und verwandte Gebärden.

¹⁴ Genaueres zu diesem Arbeitsschritt s. Langer et al. 2020 *Lemmatisierungsregeln*. (AP10-2016-01).

¹⁵ Herbst & Klotz (2003) nennen diesem Schritt *Lemmatisierung* und sprechen von der *Lemmatisierungspraxis* (vgl. 174-179), bei Atkins & Rundell (2008) ist dieser Schritt mit *Building a Headword List* überschrieben (s. 178ff.). In Anlehnung an Svenséns *establishment of lemmas* nennen wir diesen Schritt *Lemmakonstituierung* bzw. *Konstituierung der Lemmazeichen*, um ihn von der Lemmatisierung bei der Annotation zu unterscheiden. Im DGS-Korpus-Projekt sind die Vorgehensweisen bei der Annotation (Token-Type-Zuordnung) und in der Lexikographie (Lemmakonstituierung) nicht ganz deckungsgleich (vgl. Langer et al. 2016). Siehe auch Svensén (2009, 94): "...ESTABLISHMENT OF LEMMAS. This operation is not simply the same thing as lemmatization... Naturally, the establishment of lemmas assumes previous lemmatization, but it also includes deciding how lexical items having identical base forms are to be presented in the dictionary, and to what extent word elements and multi-word lexical items are to be accorded lemma status."

¹⁶ Als Proto-Einträge werden hier die schon weitgehend als Einträge strukturierten Ergebnisse der Analyse zu einem Lemmazeichen inklusive aller Notizen, Hypothesen und weiteren Informationen verstanden, wie sie Atkins und Rundell als *pre-dictionary database* für die korpusbasierte Wörterbucharbeit empfehlen. Sie enthält unter anderem auch feiner unterschiedene Proto-Lesarten (Ergebnis von *Splitting*), aus denen im Redaktionsprozess dann die eigentlichen Lesarten für das Wörterbuch z.B. durch Auswahl und ggf. auch Zusammenfassung (*Lumping*) erarbeitet werden (vgl. Atkins & Rundell 2008, z.B. 99ff., 267f.).

5. Überblick über Kollokationen und Zusammensetzungen

Kollokatoren eines Wortes können wertvolle Hinweise bei der Identifizierung seiner Bedeutungen/Lesarten liefern (vgl. Kilgariff 2012, S. 7; Kilgariff & Rundell 2003). Die ursprünglich in Sketch-Engine verwendete Formel zur Identifizierung von Kollokationen (eine Form des mutual information score oder MI-score¹⁷; vgl. Lexical Computing 2015, 2), die eine vorher annotierte grammatische Verbindung der Wörter (POS-Tagging) nicht voraussetzt, wurde mit einer kleinen Anpassung auf unsere Daten angewendet (per SQL-Abfrage in iLex beim Type-Eintrag als Listenabfrage eingebaut). So steht für alle Gebärden mit einer ausreichenden Tokenzahl eine zumindest grobe Voranalyse der häufigsten Kollokator-Kandidaten für die unmittelbaren linken und rechten Nachbarn zur Verfügung.¹⁸ Neben Kollokatoren zeigt das Ergebnis der Abfrage auch häufige LBG-artige Zusammensetzungen (mit analogem Aufbau zu deutschen Komposita). Bei der Analyse werden diese Abfrageergebnisse mit berücksichtigt. Zusammensetzungen und für den Gesamteintrag relevante, aus mehreren Gebärden bestehende Verwendungsmuster werden festgehalten,¹⁹ insbesondere die Kombinationen, die als Run-ons im Eintrag aufgeführt werden sollen.

Manche Kollokator-Kandidaten werden als „idiomatische Phrasen“ eingestuft. Das sind häufige, DGS-typische Kombinationen, deren Gesamtbedeutung auf nicht vorhersagbare Weise spezifischer ist als die Bedeutungen ihrer Teile, sowie solche, die keine Entsprechung mit parallelem Aufbau im Deutschen haben. Da sie eine eigene Bedeutungsbeschreibung erfordern oder weitere Angaben zu ihrer Rezeption oder Verwendung für den WB-Nutzer sinnvoll sein können, werden sie in den Vorabeinträgen als eigene Lesart nach allen anderen Lesarten im Wörterbucheintrag aufgeführt (vgl. Langer et al. 2019).

6. Lesarten-Differenzierung (Word Sense Discrimination WSD)

Mindestens alle Lexeme mit einer Korpus-Beleglage ≥ 25 und, soweit es auf Grundlage der Datenlage nach Ermessen der Bearbeiterin sinnvoll erscheint, auch schwächer belegte Lexeme werden dann nacheinander zur Lesartendifferenzierung bearbeitet. Die Ergebnisse werden in Proto-Lesarten in FMP dokumentiert. Bei diesem Bearbeitungsschritt werden auch feinere Unterschiede notiert (*Splitting*), die ggf. später im Produkt zu einer größeren Lesart zusammengefasst werden können (*Lumping*). Dies entspricht dem von Atkins & Rundell (2008, 97-102) beschriebenen Vorgehen und hat sich bei der Bearbeitung der ersten Einträge nach dem hier beschriebenen Verfahren als sinnvoll erwiesen.

¹⁷ Die für Sketch Engine angegebene Formel entspricht Pointwise Mutual Information. Näheres zur Verwendung dieser Formel bei uns siehe Langer & Schulder 2020.

¹⁸ Da einerseits in der Transkription schon kleine Formunterschiede (Ausführungsvarianten) durch verschiedene Subtypes bzw. Types erfasst werden und andererseits oft mehrere lexikalische Varianten für dieselbe Bedeutung existieren, ergab die Berechnung des PMI-Scores für vollständige Nachbar-Glossen nur selten Kollokator-Kandidaten, auch wenn bei der Einzeltokenanalyse bestimmte Kombinationen (eher am Glossennamen als an der vollständigen Glosse orientiert) als häufig auffielen. Bei der Berechnung müsste man Tokens von Ausführungsvarianten zusammen behandeln. Will man die Kollokationsanalyse als Hinweis auf verschiedene Lesarten der Basis nutzen, so ist es mehr oder weniger unerheblich, dass verschiedene lexikalische Varianten derselben Bedeutung als Nachbargebärde auftreten, solange das Kombinationsmuster insgesamt besonders häufig ist. Daher beziehen wir den Basis-Type nicht auf die spezifischen Types der linken und rechten Nachbarn, sondern vergrößern die Glossen der Nachbarn auf den Glossennamen. Auf diese Weise erkennen wir semantisch relevante Kombinationsmuster, auch wenn die Nachbarn verschiedenen lexikalischen oder Ausführungs-Varianten angehören. Dieses Verfahren stellt natürlich eine Vergrößerung dar und ist auch insofern ungenau, als dass der Glossenname (Lexem-Ebene) eigentlich nicht die Funktion einer Bedeutungsangabe hat. In der Praxis erweisen sich die so berechneten Angaben aber als durchaus hilfreich bei der Lesartenanalyse zur Entdeckung von relevanten Kombinationen und Mustern (s. Anhang, Folie 9; siehe auch Langer & Schulder 2020).

¹⁹ Für das produktionsgerechte Festhalten von Kollokationsmustern gibt es in FMP seit Mai 2019 entsprechende Strukturen. Siehe auch Langer et al. 2019.

Die Arbeitsschritte 6a-6c werden hier in einer logischen Reihenfolge beschrieben, finden aber in der Praxis meist gemeinsam und parallel in einem rekursiven Gesamtprozess statt.

6a. Überblick über die Beleglage und Auswahl der zu bearbeitenden Tokens

In der Listendarstellung der Tokens eines Lexems, das dem Lemmazeichen-Kandidat-Type zugeordnet ist, kann sich die Bearbeiterin einen Überblick über die Beleglage und anhand der Übersetzungen einen ersten Eindruck davon verschaffen, in welchen inhaltlichen Kontexten die Gebärde verwendet wird. Die Zuordnung von Tokens zu Lexemen stellt eine erste, wenn auch grobe Vorsortierung dar. Diese Sichtung dient dazu, sich einen Überblick über die Beleglage zu verschaffen und erste Lesartenunterscheidungen innerhalb und unterhalb der Lexem-Ebene zu erkennen. Bei der Auswahl der zu bearbeitenden Tokens achten die Bearbeiterinnen darauf, dass möglichst viele verschiedene oder zumindest eine ausgewogene Auswahl der Informanten und Regionen, sowie bevorzugt Vorkommen aus Erhebungstasks berücksichtigt werden, die den Charakter von Erzählungen oder Gesprächen haben. Außerdem sollen möglichst viele Themenbereiche abgedeckt werden, soweit dies an der Übersetzung oder am Task/Subtask erkennbar ist.

6b. Sichten und Analyse einzelner Belege

Mindestens 20, bei Bedarf aber auch mehr Belege werden im Detail angeschaut. Dazu stehen verschiedene Sichten zur Verfügung:

- A. Tokenzeile im Tokentab des Types (s. Anhang Folie 12) mit nach Bedarf der Bearbeiterinnen weiteren eingeblendeten Informationen zum Token (z.B. Übersetzung, linker und rechter Nachbar), oder/und zum Bearbeitungsstand (z.B. Bearbeiter-Notizen, zugeordnete Lesartennummer, Workflow-Angaben),
- B. vollständiges Transkript mit unmittelbarem Zugriff auf die Originalfilme (s. Anhang Folie 13),
- C. Tokens-im-Kontext-Sicht – mit der Möglichkeit, den Filmausschnitt gezielt anzuschauen (s. Anhang Folie 14)²⁰.

Die Bearbeiterin schaut sich immer die Originaldaten im Film an und macht sich Notizen zu der Bedeutung der Gebärde in diesem konkreten Kontext, zum Kontext selbst und zu allen weiteren Auffälligkeiten, die bemerkt werden. Dabei kann es sich u.a. um eine auffällige Ausführung, eine spezielle Wortform oder eine Einbettung in eine Constructed Action (Rollenübernahme) handeln. Bei anderen Gebärden ist es vielleicht relevant zu untersuchen, welche Aktanten beteiligt sind und ob die Gebärde ggf. räumlich auf diese Bezug nimmt oder Ähnliches. Dazu muss die Bearbeiterin sich die gebärdete Äußerung ansehen und ggf. auch den weiteren vorhergehenden Kontext sichten (im Transkript B), um die Bezüge auflösen oder die Rollen identifizieren zu können. Für die persönlichen Bearbeiternotizen können in iLex mehrere Metadatenkategorien am Token-Tag genutzt werden, die bei der Bearbeitung je nach Bedarf in einer entsprechenden Tokenlistenansicht (A) angezeigt werden.

Wenn sich bei diesem Schritt bereits einzelne Lesarten herauskristallisieren, werden diese gleich im DWS angelegt und zum Taggen verwendet (siehe unten Arbeitsschritt 6c).

Wurden für alle relevanten Lexeme entsprechende Tokens bearbeitet, werden auch die Tokens, die keinem Lexem, sondern direkt der Gebärde zugeordnet sind, gesichtet, um herauszufinden, ob sich darin weitere, bisher nicht identifizierte Lesarten finden lassen. Dies geschieht u.a. auch mit Blick auf auffällige Mundbild-Häufungen.

²⁰ In der Praxis wurde anfänglich viel mit dieser Ansicht (C) experimentiert und gearbeitet. Mittlerweile bevorzugen die Bearbeiterinnen eher die Ansicht der Tokenzeile (A) mit je nach Blickwinkel und Fragestellung individuell angepassten unterschiedlichen Abfragen im Wechsel, sowie das Transkript (B).

6c. Gruppierung der Belege und Anlage der Proto-Lesarten

Bei der Sichtung und anschließender Zusammenschau der einzelnen Belege kristallisieren sich gleichartige Verwendungen heraus, die in FMP als Proto-Lesarten angelegt werden.

FMP vergibt für neu angelegte Lesarten automatisch eindeutige IDs, mit denen entsprechende Vorkommen in iLex dann getaggt werden können (sense-tagging). Umgekehrt können gute Beispiele für die Lesarten per Drag & Drop des Tokentag-Ankers in FMP gezogen werden, wo sie dann für die Beispielauswahl (siehe Abschnitt 9) zur Verfügung stehen (s. Anhang Folie 17).

Zu jeder Lesart werden, soweit vorhanden oder von der Bearbeiterin erstellbar, folgende Beobachtungen und Informationen entweder in internen Kommentarfeldern oder in entsprechenden produktionsrelevanten Feldern festgehalten:

- Bedeutung/Verwendungsbedingungen (lexikographische Definition): Es gibt ein Feld für eine interne Definition und ein Feld für die Formulierung, wie sie im Produkt erscheinen soll;
- geeignete Beleg-Tokens werden per drag & drop aus iLex in FMP gezogen; aus diesen Belegen für eine Lesart werden besonders geeignete als redaktionelle Beispiele für das Wörterbuch ausgesucht und in FMP und iLex zur Weiterbearbeitung markiert;
- typische Mundbilder;
- geeignete Übersetzungsäquivalente, ggf. auch geeignete disambiguierende Zusätze zu den Übersetzungsäquivalenten;
- synonyme/teilsynonyme oder antonyme Gebärden – soweit bekannt oder im gebärdeten Kontext vorhanden: per drag & drop des Type-Ankers aus iLex in FMP;
- aus den Zusammensetzungen (Kollokator-Kandidaten – Abfrage über die Daten) werden relevante Kollokationen und Kollokationsmuster, die zu der jeweiligen Lesart gehören, in FM als „Muster“ angelegt und der Lesart zugeordnet;
- semantische oder/und syntaktische Auffälligkeiten, grammatisches Verhalten;
- alle weiteren Beobachtungen, Auffälligkeiten, Fragen und Hypothesen zur Lesart werden in internen Kommentarfeldern festgehalten;
- Sachgruppenzuordnung: Einer Lesart werden bis zu drei Sachgruppen zugeordnet.²¹

6d. Einstufung der Lesarten

Die Proto-Lesarten werden je nach Datenlage als für die Produktion relevant markiert und, falls sinnvoll, zu größeren Lesarten zusammengefasst (*Lumping*: siehe auch Abschnitt 8 unten) oder als nicht zu produzieren gekennzeichnet und ggf. unter Beobachtung gestellt.

7. Klärung offener Fragen

Sofern sich aus der Bearbeitung grundsätzliche oder singuläre Fragen ergeben, werden diese nach Möglichkeit geklärt. Dazu gibt es verschiedene Möglichkeiten, die im Folgenden genannt werden (siehe auch Fußnote 1):

7a. Detailtranskription beauftragen

Wenn sich aus der Sichtung oder der Bearbeitung der Einzelbelege weitere Fragen an die Daten ergeben, so können entsprechende Bearbeitungsaufträge an die Detailtranskription gegeben werden.

²¹ Eine Sachgruppenzuordnung fällt bei der Bearbeitung der einzelnen Lesart vergleichsweise leicht, da man sich ohnehin mit der Bedeutung und Verwendung intensiv beschäftigt. Die Sachgruppenzuordnung wird zunächst relativ intuitiv getroffen. FMP stellt bereits angelegte Sachgruppen als Auswahlliste zur Verfügung. Zu einem geeigneten Zeitpunkt werden die Sachgruppen überarbeitet und vereinheitlicht.

7b. Fokusgruppe

Die Fokusgruppe steht für Nachfragen zu einzelnen Gebärden zur Verfügung, insbesondere wenn es sich um regionale Gebärden handelt.

7c. DGS-Feedback

Das DGS-Feedback steht zwar für bestimmte Fragestellungen prinzipiell zur Verfügung, jedoch sind die Fragestellungen, die sich aus der Bearbeitung der Lesarten ergeben, oft zu speziell, als dass sie über das relativ grobe Instrument des DGS-Feedbacks geklärt werden könnten. Es eignet sich vor allem zur ergänzenden Erhebung von regionalen Gebärden für semantische Sets (z.B. Farben, Wochentage und ähnliches; Fragetyp 2), die keine allzu feine Kontextualisierung der jeweils gemeinten Lesarten benötigen.

7d. Nacherhebungen

Auch über die Fokusgruppe hinaus können ggf. kleinere Nacherhebungen gemacht werden. So wurden z.B. 2018 Gebärden für Städtenamen mit Hilfe des neu entwickelten Erhebungstools SignHunter auf den 6. Deutschen Kulturtagen in Potsdam (s. Hanke et al. 2020) erhoben.

7e. Redaktionssitzung

In der Redaktionssitzung werden gemeinsam die Fragen besprochen, die eine einzelne Bearbeiterin nicht allein entscheiden kann oder die der gemeinsamen Diskussion bedürfen oder die grundsätzlich entschieden werden müssen.

Artikelschreibung

Ist ein Proto-Eintrag soweit bearbeitet, dass er bei der nächsten Produktion mit exportiert und produziert werden soll, wird er entsprechend gekennzeichnet und durchläuft eine Reihe von Checks zur Vollständigkeit und Konsistenz der Angaben. In FMP sind etliche Checks eingebaut, die automatisch fehlende Angaben oder einige Widersprüche melden. Die folgenden Arbeitsschritte werden spätestens zu diesem Zeitpunkt erledigt, wenn sie bis zu diesem Zeitpunkt noch nicht abgeschlossen sind. Häufig wird in der Praxis jedoch ein Großteil der Arbeitsschritte 8-10 zeitgleich und im ständigen Wechsel ausgeführt.

8. Auswahl und Zusammenfassung der Lesarten fürs Produkt

Bei der Artikelschreibung werden feinere Unterscheidungen dort, wo es aus Produkt- und Nutzersicht sinnvoll erscheint, zu größeren Lesarten zusammengefasst (*Lumping*) und für das Produkt ausgewählt.²² Folgendes wird entschieden und entsprechend markiert oder umgesetzt:

- Welche Proto-Lesarten kommen ins Produkt (gesicherte Erkenntnisse)?
- Welche Proto-Lesarten werden zu einer Lesart zusammengefasst? Diese größeren Lesarten werden angelegt, die entsprechenden Angaben, insbesondere die Definitionen, Übersetzungsäquivalente und die Beispielauswahl werden angepasst.
- Welche Proto-Lesarten stehen unter Beobachtung (z.B. weil die Datenlage noch zu dünn ist)?

²² Dies entspricht dem von Atkins & Rundell beschriebenen Vorgehen (2008, 97-102). Wir sind mit Atkins & Rundell (vgl. S. 99-100) der Auffassung, dass das Ergebnis der Analyse weiterhin dokumentiert und erhalten bleiben sollte, auch wenn Lesarten für das Produkt zu größeren Einheiten zusammengefasst werden. Strukturen, die verschiedene Detaillierungsgrade der Lesarten darstellen können, sind für das DWS angedacht, aber in FMP nur behelfsmäßig realisiert. Zurzeit arbeiten wir in FMP mit einer Workaround-Lösung von strukturell gleichwertigen Lesarten und einer eher groben Lumping-Möglichkeit über Verweise zwischen mehreren Lesarten eines Eintrags.

- Welche Proto-Lesarten fallen für das Produkt weg (z.B. weil die Lesartenunterscheidung für das Produkt zu kleinteilig oder marginal oder noch nicht ausreichend abgesichert ist)?

9. Korpusbeispiele für die Produktion auswählen und bearbeiten

Für jede der für die Produktion vorgesehenen Lesarten werden aus den bei der Lesart vermerkten Korpusbeispielen normalerweise zwei ausgewählt, die im Eintrag als Veranschaulichung bei der jeweiligen Lesart gezeigt werden. Diese müssen für die Produktion entsprechend aufbereitet werden (s. Anhang Folie 17, Langer et al. 2018b). Die Aufbereitung beinhaltet:

- Den als Beispiel gezeigten Ausschnitt des Films festlegen: Da im Zuge der Basistranskription keine DGS-Satzgrenzen bestimmt werden und beim jetzigen Forschungsstand auch oft gar nicht bestimmt werden können, gibt es im Transkript eine Einteilung in kürzere Sinnabschnitte nur durch die Alignierung der deutschen Übersetzung. Diese Abschnitte stimmen oft, aber eben nicht immer, mit den Strukturgrenzen einer DGS-Äußerung überein. Daher wird bei den ausgesuchten Beispielen geprüft, ob unter Berücksichtigung erkennbarer DGS-Strukturgrenzen ein anderer, besser geeigneter Ausschnitt für das Beispiel bestimmt werden soll. Ggf. werden auch mehrere kurze Äußerungen zusammen ausgewählt, um einen entsprechenden semantischen Kontext für die Zielgebärde im Beispiel zugänglich zu machen.²³
- Anpassung der Übersetzung: Je nach gewähltem Ausschnitt muss ggf. die deutsche Übersetzung für das Beispiel angepasst werden, damit sie dem gewählten Ausschnitt auch entspricht. Darüber hinaus sind ggf. weitere Anpassungen der Übersetzung nötig, wenn sie z.B. zu frei ist und dadurch genau die Gebärde, um die es im Beispiel geht, nicht explizit berücksichtigt oder Ähnliches. Die angepasste Übersetzung wird zusätzlich in iLex gespeichert und ersetzt nicht die für das Teilkorpus angefertigte. In der angepassten Übersetzung wird der Teil der Übersetzung fett markiert, der der Zielgebärde entspricht.
- Hinweis auf den Kontext erstellen: Manche Beispielausschnitte sind aus dem Kontext gerissen nicht oder nicht leicht verständlich, z.B. weil das Thema oder vorher eingeführte Personen, Orte und Bezüge nicht bekannt sind. Viele Beispiele erhalten daher eine kurze einführende Bemerkung zum Kontext, aus dem das Beispiel stammt. Das kann ein sehr kurzer Hinweis auf das Gesprächsthema sein bis hin zu einer kurzen Zusammenfassung der vorangegangenen Inhalte mit Benennung der Akteure und Bezüge.²⁴

10. Alle Angaben vervollständigen und in Form bringen

Alle Angaben zur Lesart werden im Zusammenhang geprüft, ergänzt und entsprechend den bereits festgelegten Vorgaben (Style Manual²⁵) formuliert. Spätestens jetzt wird die Definition für das Produkt ausformuliert.

Spätestens jetzt muss außerdem festgelegt werden:

- Welche Variante ist die Hauptvariante, die das Lemma insgesamt repräsentiert?

²³ In der Diskussion beim Fokusgruppentreffen 2016 hat sich herausgestellt, dass Gehörlose häufig mit sehr viel kürzeren Kontext-Ausschnitten zurechtkommen als zunächst vermutet. Da das Wörterbuch aber verschiedenen Nutzergruppen dient, wird ggf. trotzdem ein etwas größerer Ausschnitt gewählt.

²⁴ Mit Blick auf die Nutzergruppe DGS-Lerner tendieren wir dazu, Kontexte häufiger anzugeben, als dies von muttersprachlichen Nutzern vielleicht benötigt würde.

²⁵ Die erarbeiteten und jeweils Vorgaben für die einzelnen Elemente sind im Mitarbeiter-Wiki in verschiedenen Seiten festgehalten und beschrieben. Das Style-Manual existiert also nicht als ein einzelnes durchgehendes Dokument.

- In welcher Reihenfolge sollen die Lesarten im Produkt erscheinen? Die Reihenfolge im Produkt folgt grob der Vorkommenshäufigkeit im Korpus, wie sie sich aus der Bearbeitung der Einträge aufdrängt, ohne dass dies endgültig quantifiziert werden kann, da oft nicht alle Vorkommen durchgesehen und bearbeitet werden und nicht alle bearbeiteten Tokens immer eindeutig genau einer Lesart zugeordnet werden können. Darüber hinaus wird versucht, eine intuitiv sinnige Reihenfolge der Lesarten auf Grundlage der semantischen Nähe zu erreichen.
- Reihenfolge der Übersetzungsäquivalente für jede Lesart festlegen und entscheiden, welche Übersetzungsäquivalente auch im Deutschindex erscheinen sollen und welche nur im Eintrag.
- Endkontrolle der Beispiele: sind alle ausgewählten Beispiele bearbeitet worden und in Ordnung?

Produktion

Wenn die Artikelschreibung abgeschlossen ist, steht fest, welche Studioaufnahmen für den Eintrag benötigt werden. Die folgenden Arbeitsschritte sind nur kurz aufgelistet, um den Workflow zu veranschaulichen, aber nicht im Detail beschrieben.

11. Videoproduktion

11a. Regielistenerstellung

Die Regieliste wird für die ausgewählten, zu produzierenden Einträge aus FMP exportiert und weiterverarbeitet. Ist die Glossenliste und die Liste bereits aufgenommener, lemmatisierter und geprüfter Studiofilme in FMP aktuell, werden per Script nur diejenigen benötigten Angaben in der Regieliste versammelt, die noch fehlen.

Für die Produktion im Studio und die anschließende Lemmatisierung wird aus dem Export eine Protokollliste erstellt. Darüber hinaus wird zur Unterstützung der Gebärdendarsteller_innen ein PDF-Dokument erzeugt, das die zu gebärdenden Formen (mit HamNoSys und Mundbild) auf dem Monitor anzeigt (s. Anhang Folie 18).

11b. Studioaufnahmen

Für alle als Film zu realisierenden Formangaben und Verweise werden saubere Studioaufnahmen der isolierten Gebärdenform benötigt. Das sind im Einzelnen:

- Lemma (Zitatform des Lemmazeichens), ohne Mundbild;
- Ausführungsvarianten des Lemmas, ohne Mundbild;
- Zitatformen der Lemmazeichen, auf die verwiesen wird (formähnliche und verwandte Gebärden, Synonyme, Antonyme), einmal ohne Mundbild für die Eintragsrepräsentation und einmal mit lesartenspezifischem Mundbild für die lesartenspezifischen Verweise.

Dazu kommen noch Studioaufnahmen für:

- idiomatische Phrasen: jeweils eine der möglichen Kombination wird als repräsentativer Film für das gesamte Muster ausgewählt und aufgenommen (ggf. mit Mundbildern oder Mundgestiken);
- Aufnahmen für alle Bestandteile der Gebärdenmuster (Kollokationen), idiomatischen Phrasen und komposita-ähnlichen Zusammensetzungen (Kombinationen aus Basis und einer Gruppe von Kollokatoren) als isolierte Einzelgebärden (ohne Mundbild).

Die Gebärden werden im Studio aus vier Perspektiven (frontal, 45°, 90° und Vogelperspektive) mit sieben Kameras (davon drei 3D-Kameras, drei 4K-Kameras und eine 2K-Kamera) aufgenommen.²⁶

11b. Postproduktion der Studioaufnahmen

Die Postproduktion der Studioaufnahmen beinhaltet:

- Anbindung der Filme an iLex
- Komprimierung der Filme
- Synchronisierung der verschiedenen Perspektiven
- Anlegen des Transkriptes
- Segmentierung und Lemmatisierung der aufgenommenen Einzelgebärden in iLex
- inhaltliche und technische Kontrolle der Aufnahmen durch eine/n gehörlosen Mitarbeiter_in
- Bestimmung eines Standbildes für jede Einzelgebärde für das Micon (in iLex)
- Bestimmung der richtigen Parameter für das Chroma-Keyen (Entfernen des Hintergrunds für die Micon-Erstellung)

12. Produktion der Vorabeinträge

Die ausgewählten zu produzierenden Einträge werden aus FMP mit allen erforderlichen Angaben als ein HTML-Dokument exportiert, das dann in der Produktion durch Skripte weiterverarbeitet wird. Die aus FMP exportierten Angaben enthalten die IDs der beteiligten Types und Beispieltokentags, über die die entsprechenden Types und zugehörigen Studio-Filme (Einzelgebärdenfilme) sowie die Beispieltags und die benötigten Beispielausschnitte identifiziert werden. Bei der Produktion wird eine entsprechende Filmproduktion angestoßen, bei der die benötigten Micons, Einzelgebärden- und Phrasenfilme sowie Korpusbeispielfilme produziert werden.

Der Export wird per Skript so umgewandelt, dass die fertigen Einträge (html-Seiten) mit eingebundenen Filmen, Standbildern und Micons entstehen und aus den Daten die verschiedenen Indexe (Deutsch-Index, Sachgruppenindex, Graph) erzeugt werden. Noch nicht bearbeitete Einträge werden als Rumpfeinträge automatisch erzeugt, wenn aus schon bearbeiteten Einträgen auf sie verwiesen wird.

Zusammen mit den Umtexten und Anhängen bilden die Filme, Standbilder und HTML-Seiten eine in sich vollständige, stand-alone benutzbare Vorabversion des Wörterbuchs, die darüber hinaus mit dem Öffentlichen DGS-Korpus verlinkt ist (s. Müller et al. 2020).

References

- Atkins, B.T. Sue / Rundell, Michael (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Hanke, Thomas / Jahn, Elena / Wähl, Sabrina / Böse, Oliver / König, Lutz (2020): SignHunter – A Sign Elicitation Tool Suitable for Deaf Events. *Community, Technological Challenges and Application Perspectives. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages. 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 2020. Paris, France: European Language Resources Association (ELRA), S. 83–88. [Online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20030.html>.]

²⁶ Stand März 2020. Geplant ist, zukünftig mit sechs 4K-Kameras aufzunehmen.

- Heid, Ulrich (2008): Corpus linguistics and lexicography. In: Lüdeling, Anke / Kytö, Merja / McEnery, Tony (eds.): *Corpus Linguistics. An international Handbook*. Berlin: Mouton de Gruyter, S. 131–153.
- Herbst, Thomas / Klotz, Michael (2003): *Lexikografie*. Paderborn: Schöningh.
- Kilgarriff, Adam / Rundell, Michael (2003): Lexical profiling software and its lexicographic applications: a case study. In: Braasch, Anna / Povlsen, Claus (eds.): *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002: Copenhagen, Denmark, August 13-17, 2002*, vol. 2, S. 807-818. [Online: http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202002/090_2002_V2_Adam%20Kilgarriff%20&%20Michael%20Rundell_Lexical%20Profiling%20Software%20and%20its%20Lexicographic%20Ap.pdf.]
- Kilgarriff, Adam (2012): *Using corpora [and the web] as data sources for dictionaries*. [Online: https://www.sketchengine.co.uk/wp-content/uploads/Using_corpora_2012.pdf.] [Gedruckte Fassung (2013) in: *The Bloomsbury Companion to Lexicography*, Howard Jackson (ed.), Bloomsbury, London. Chapter 4.1, S. 77–96.]
- Langer, Gabriele / Schulder, Marc (2020): “Collocations in Sign Language Lexicography: Towards Semantic Abstractions for Word Sense Discrimination”. In Efthimiou, Eleni et al. (eds.): *Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages. 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 2020. Paris, France: European Language Resources Association (ELRA), S. 127–134. [Online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20017.html>.]
- Langer, Gabriele / Müller, Anke / Wähl, Sabrina / Hanke, Thomas (2019): “The DGS-Korpus approach to including frequent sign combinations in a corpus-based electronic sign language dictionary”. Poster presented at the Theoretical Issues in Sign Language Research (TISLR) 13 Conference, Sep 26-28, 2019 in Hamburg, Germany. [Poster: <https://doi.org/10.25592/uhhfdm.8351>]
- Langer, Gabriele / Müller, Anke / Wähl, Sabrina (2018a): “Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS)”. In: Bono, Mayumi et al. (eds.): *Workshop Proceedings. 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan, 12 May 2018. ELRA. S. 107–114. [Online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/18026.html>.]
- Langer, Gabriele / Müller, Anke / Wähl, Sabrina / Bleicken, Julian (2018b): “Authentic Examples in a Corpus-Based Sign Language Dictionary – Why and How”. In Čibej, Jaka, Gorjanc, Vojko, Kosem, Iztok & Krek, Simon (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts, S. 483–497. [Online: <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/2933-1.pdf>.]
- Langer, Gabriele / Müller, Anke / Wähl, Sabrina / König, Susanne / Hanke, Thomas / Konrad, Reiner. (2020, October 30). *Lemmatisierungsregeln (Vorläufige Wörterbuch-Einträge) (Version 4)*. (DGS-Korpus-Arbeitspapier AP10-2016-01). [Online: <http://doi.org/10.25592/uhhfdm.1899>.]
- Langer, Gabriele / Troelsgård, Thomas / Kristoffersen, Jette / Konrad, Reiner / Hanke, Thomas / König, Susanne (2016): “Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing”. In: Efthimiou, Eleni / Stavroula-Fotinea, Evita / Hanke, Thomas / Hochgesang, Julie / Kristoffersen, Jette / Mesch, Johanna (eds.): *Workshop Proceedings. 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia, 28 May 2016. Paris: ELRA, S. 143–152. [Online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/16014.html>.]

- Lexical Computing Ltd. (2015): *Statistics used in the Sketch Engine*. July 8, 2015. [Online: <https://www.sketchengine.co.uk/wp-content/uploads/ske-statistics.pdf>.]
- Müller, Anke / Hanke, Thomas / Konrad, Reiner / Langer, Gabriele / Wähl, Sabrina (2020): “From Dictionary to Corpus and Back Again – Linking Heterogeneous Language Resources for DGS”. In Efthimiou, Eleni et al. (eds.): *Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages. 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 2020. Paris, France: European Language Resources Association (ELRA), S. 157–164. [Online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20025.html>.]
- Sinclair, John (2004): *Developing Linguistic Corpora: a Guide to Good Practice. Corpus and Text – Basic Principles*. [Online: <http://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm>, Gedruckte Fassung in: Wynne, Martin (ed.) (2004): *Developing Linguistic Corpora. A Guide to Good Practice*. Oxford: Oxford Books.]
- Svensén, Bo (2009): *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Wähl, Sabrina / Langer, Gabriele / Müller, Anke (2018): “Hand in Hand - Using Data from an Online Survey System to Support Lexicographic Work”. In: Bono, Mayumi et al. (eds.): *Workshop Proceedings. 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan, 12 May 2018. ELRA. S. 199–206. [Online: <https://www.sign-lang.uni-hamburg.de/lrec/pub/18025.html>.]

Anhang

Der folgende Anhang soll einen Eindruck von den Arbeitsumgebungen iLex und FMP vermitteln und veranschaulichen, welche Sichten auf die Daten zur Analyse und Vorbereitung der Artikelschreibung wir nutzen. Die Darstellung folgt in etwa den Bearbeitungsschritten (Workflow) anhand einiger exemplarischer Beispiele.

Arbeitsumgebung - iLex

immer vor mandarinfarbenem Hintergrund

... – Filme – Transkripte – Tokens – Types und Subtypes – Übersetzungen – Metadaten – ...

Type-Liste

Transkript mit Annotationen

Glosse	Token	HamNoSys
ZAUBEREI1-\$\$AM	1	...
ZAUBEREI2-\$\$AM	1	...
ZAUBEREI3-\$\$AM	12	...
ZEBRA1-\$\$AM	3	...
ZEICHEN1-\$\$AM	85	...
ZEICHEN2-\$\$AM	28	...
ZEICHEN3-\$\$AM	50	...
ZEIGEN-PAPIER11.XSAM	1	...
ZEIGEN-PAPIER2-\$\$AM	3	...
ZEIGEN1A-\$\$AM	235	...
ZEIGEN1B-\$\$AM	28	...
ZEIGER1-\$\$AM	73	...
ZEIT1-\$\$AM	550	...
ZEIT3-\$\$AM	18	...
ZEIT5A-\$\$AM	130	...
ZEIT5B-\$\$AM	49	...
ZEIT6-\$\$AM	12	...
ZEIT7A-\$\$AM	84	...
ZEIT7B-\$\$AM	27	...
ZEIT7C-\$\$AM	4	...
ZEIT9-\$\$AM	17	...
ZEITUNG1A-\$\$AM	64	...
ZEITUNG1B-\$\$AM	30	...
ZELT1-\$\$AM	29	...
ZERFALLEN1-\$\$AM	2	...
ZERFRESSEN1-\$\$AM	7	...
ZERFRESSEN2-\$\$AM	2	...
ZERKLEINERUNG2-\$\$AM	3	...
ZERSTÖREN2-\$\$AM	10	...
ZEUGE1-\$\$AM	32	...
ZEUGE2-\$\$AM	123	...
ZEUGE3-\$\$AM	24	...
ZEUGE6-\$\$AM	5	...
ZEUGE7-\$\$AM	1	...
ZEUGNIS1-\$\$AM	6	...

Timecode	Deutsche Über...	Englische Über...	Segmentierung...	Lexem/Gebärd...	HamNoSys_A	HamNoSys_Ab...	Mundbild/Mun...	Beispiel_A
11:03:48:06			x	KOMMEN4 phs:2	...	C		
11:03:48:19								
11:03:48:25	ich hatte aber keine Zeit, ich hatte eine eigige...	I didn't have any time, though I had...						ich hatte aber keine Zeit, ich habe Familie. Das hat er auch verstanden.
11:03:48:26			x	ICH1'ssam	...	kein		
11:03:48:31								
11:03:48:32			x	SOEST-OFF-\$\$AM	...			
11:03:48:33								
11:03:48:38			x	ICH1	...	h		
11:03:48:38								
11:03:48:49			x	ZEIT1	...	zeit		
11:03:49:10								
11:03:49:20			x	familie		
11:03:49:25								
11:03:49:38			x	VERSTEHEN1 phs:2	...	verstehen		
11:03:50:07								
11:03:50:07			x	AUF-PERSON1 phs:1	...			
11:03:50:15								
11:03:50:23			x	wie		
11:03:50:23								
11:03:50:30			x	SOEST-OFF-\$\$AM	...			
11:03:50:41								
11:03:50:44			x	wie		
11:03:51:00								
11:03:51:09			x	SPEZIAL-WIE-WIE	...	wie		
11:03:51:17								
11:03:51:20			x	WENN1	...	wenn		
11:03:51:32								
11:03:51:44			x	SOLO1 ph:1	...	solo		
11:03:51:49								
11:03:52:05			x	ICH1	...			

Type-Eintrag

Token-Tag

Gebärden: ZEIT1-\$\$AM

Form Kinder QxQ Tokens Analyse Sprache Voten Standb. Profil +

Glosse: ZEIT1-\$\$AM

Glosse englisch: TIME

HamNoSys: [Handwritten HamNoSys code]

Mund: [Handwritten mouth image]

Zusammengesetzt: [Handwritten code]

Beschreibung: [Handwritten description]

Abgeleitet von: [Handwritten code]

Glosse: [Handwritten code]

2003-09-08 (Annika Fie...

11:03:48:49: ZEIT1

Tokens Analyse Standbilder Feedback-Verw. Profil +

Timecode: 11:03:48:49 - 11:03:49:10

Art Wert Wer?

GL_Lesart_vor... 10.* Zeit als Ressource ... GL

WB: Kontext ... Jemand wollte den Sprec... GL

WB: LA in Fil... 2313 GL

WB: Workflow... ✓ GL_ok GL

WB: Workflow... Schnitt: ✓ | Übers: ✓ | Ko... GL

Glosse: ZEIT1

HamNoSys: [Handwritten HamNoSys code]

mod: [Empty field]

HamNoSys: [Empty field]

dir: [Empty field]

lok: [Empty field]

Bedeutung: [Empty field]

Pluralität: Nicht zutreffend

Interpretation: [Empty field]

2014-08-28 (Ilona Hofmann)

Subtypes

Tokens

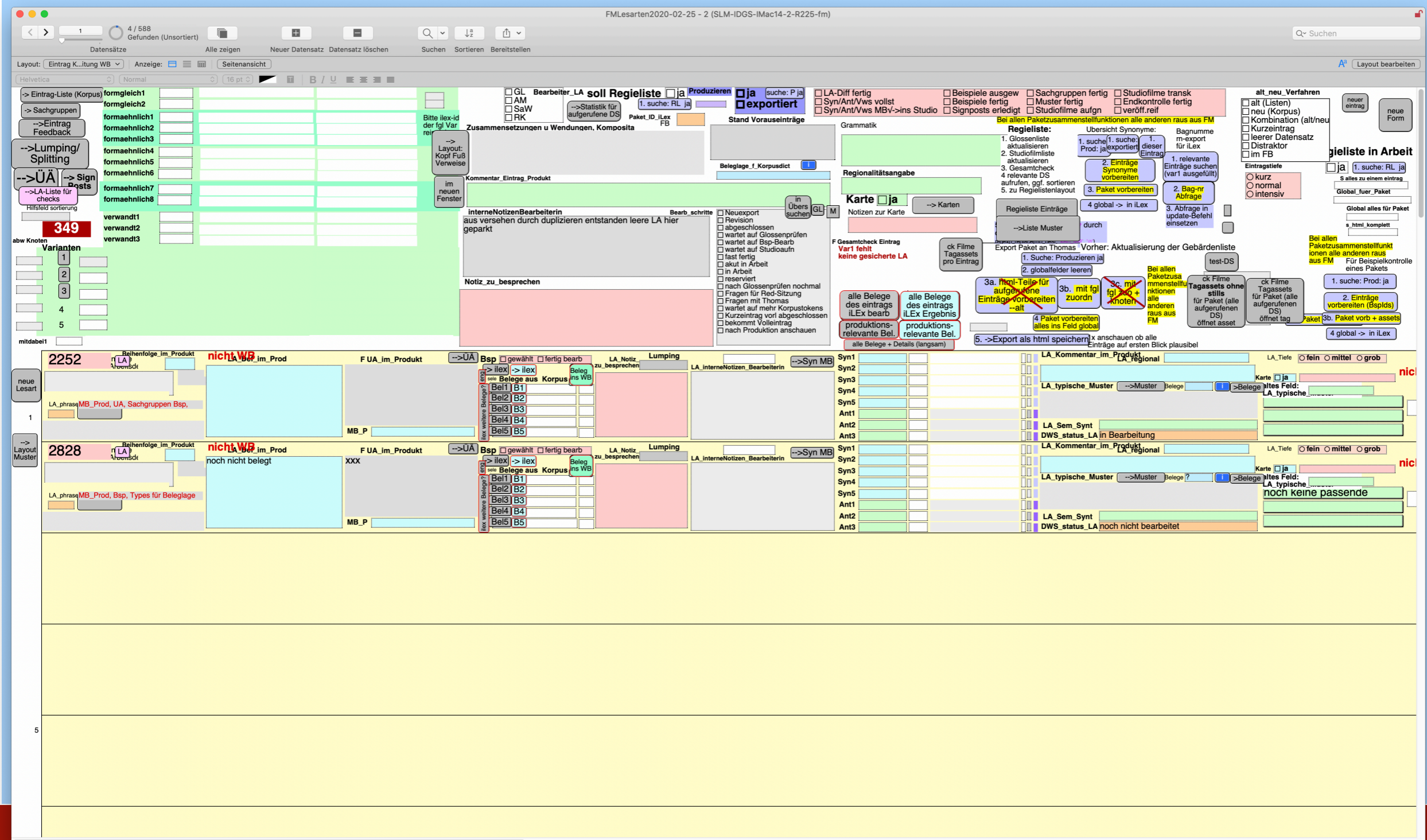
Arbeitsumgebung - Proto-Eintrag in FileMaker

immer vor hellblauem Hintergrund

- A - Angaben zum gesamten Lemmazeichen/Eintrag
- B - Lesarten und Angaben zu einzelnen Lesarten

A

B



Sicht: Gebärdenliste

Korpus-Gebärden

zeit

23 Einträge

Glosse	Token	HamNoSys
\$MAN-ZEITRAUM1	1	⊕ ₁₀ ↓ ₁₀ ⊗ ₁₀
\$MAN-ZEITRAUM2	1	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
\$NUM-ZEIT-NACH-JETZT1-\$SAM	48	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
\$NUM-ZEIT-VOR-JETZT1-\$SAM	89	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
\$UNBEKANNT_NUE18_FREIZEIT-\$SAM	1	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
JAHRESZEIT1-\$SAM	5	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LANG-ZEIT1A-\$SAM	84	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LANG-ZEIT1B-\$SAM	16	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LANG-ZEIT1C-\$SAM	35	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LANG-ZEIT1D-\$SAM	4	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LANG-ZEIT2-\$SAM	33	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LANG-ZEIT3-\$SAM	3	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
LESEN-ZEITUNG1-\$SAM	7	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
ZEIT1-\$SAM	522	⊕₁₀⊗₁₀⊕₁₀⊗₁₀
ZEIT3-\$SAM	26	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
ZEIT5-\$SAM	210	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
ZEIT6-\$SAM	14	⊕₁₀⊗₁₀⊕₁₀⊗₁₀
ZEIT7A-\$SAM	85	⊕₁₀⊗₁₀⊕₁₀⊗₁₀
ZEIT7B-\$SAM	19	⊕₁₀⊗₁₀⊕₁₀⊗₁₀
ZEIT7C-\$SAM	4	⊕₁₀⊗₁₀⊕₁₀⊗₁₀
ZEIT9-\$SAM	13	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
ZEITUNG1A-\$SAM	59	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀
ZEITUNG1B-\$SAM	27	⊕ ₁₀ ⊗ ₁₀ ⊕ ₁₀ ⊗ ₁₀

Beispiel:

ZEIT1A-\$SAM [⊕₁₀⊗₁₀⊕₁₀⊗₁₀](⊕₁₀⊗₁₀)

ZEIT1-\$SAM und ZEIT6-\$SAM:

formähnlich → Verdacht auf Varianten

(geprüft: ✓)

Varianten

1 ilex://types.id=16979

ZEIT1-\$SAM

[⊕₁₀⊗₁₀⊕₁₀⊗₁₀](⊕₁₀⊗₁₀)

2 ilex://types.id=45276

ZEIT6-\$SAM

[⊕₁₀⊗₁₀⊕₁₀⊗₁₀](⊕₁₀⊗₁₀)

Beispiel:

ZEIT7A-\$SAM ⊕₁₀⊗₁₀⊕₁₀⊗₁₀

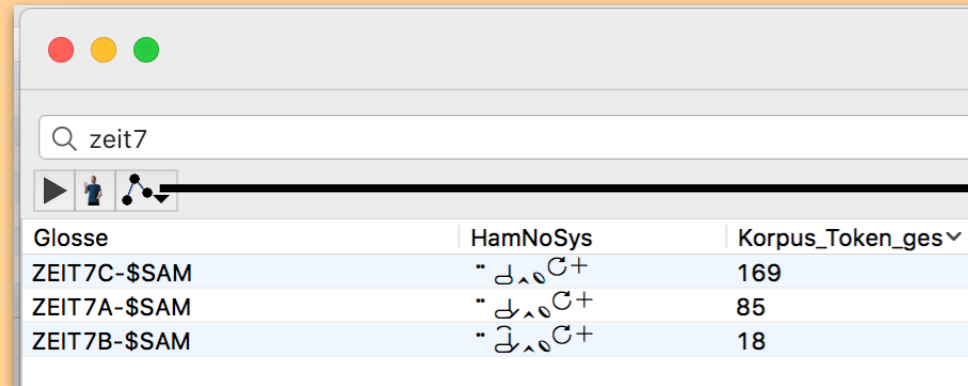
→ vermutlich Varianten-Kandidaten?

ZEIT7B-\$SAM ⊕₁₀⊗₁₀⊕₁₀⊗₁₀

ZEIT7C-\$SAM ⊕₁₀⊗₁₀⊕₁₀⊗₁₀

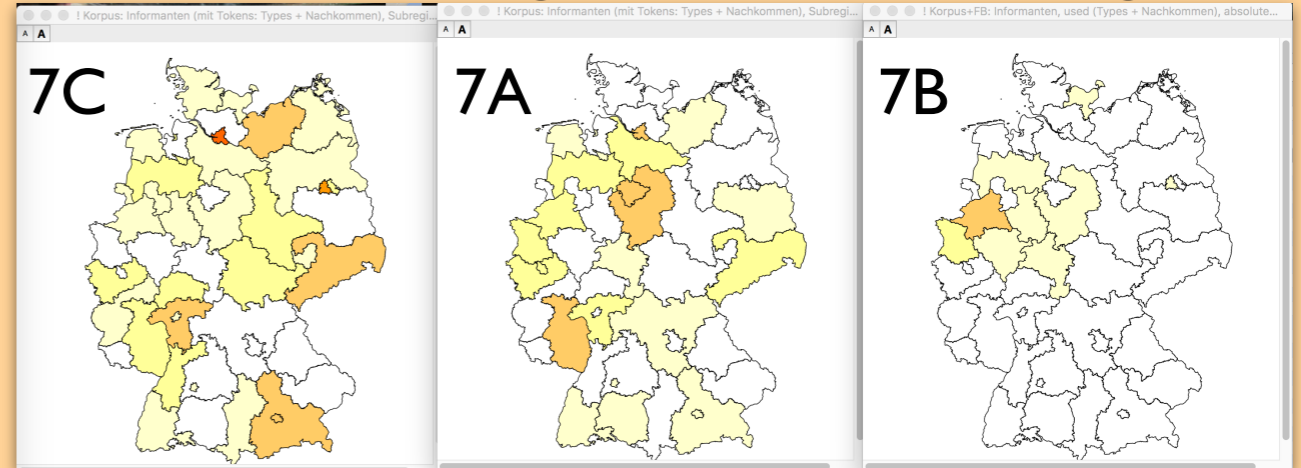
Beispiel: ZEIT7A/B/C-\$SAM

Sicht: Gebärdenliste



Glosse	HamNoSys	Korpus-Token_ges
ZEIT7C-\$SAM	" ɔ̯ ʌ 0 ʘ +	169
ZEIT7A-\$SAM	" ɔ̯ ʌ 0 ʘ +	85
ZEIT7B-\$SAM	" ɔ̯ ʌ 0 ʘ +	18

Karten zur regionalen Verteilung



→ 3 Ausführungsvarianten

→ 7B: regionale Variante

Varianten

1 ilex://types.id=13746

ZEIT7C-\$SAM

" ɔ̯ ʌ 0 ʘ +

2 ilex://types.id=14984

ZEIT7A-\$SAM

" ɔ̯ ʌ 0 ʘ +

3 ilex://types.id=14983

ZEIT7B-\$SAM

" ɔ̯ ʌ 0 ʘ +

Regionalitätsangabe

Form 360.3 hauptsächlich in Nordrhein-Westfalen

Detailansicht Daten: Händigkeit

Gebärden: ZEIT7C-\$SAM

Form	Kinder	QxQ	Tokens	~	∞
11 Einträge					
lexem	hns	Korpus	Tokens_k	1-Hd	2-Hd
FREIZEIT1A	" ↘ ↗ ↻ +	Korpus: 2	2	0	2
FREIZEIT1B	" ↘ ↗ ↻ +	Korpus: 21	21	0	21
FREIZEIT1C	" ↘ ↗ ↻ +	Korpus: 4	4	0	4
STUNDE1A	↘ ↗ ↻	Korpus: 35	35	34	1
ZEIT7A	" ↘ ↗ ↻ +	Korpus: 63	63	14	49
ZEIT7B	" ↘ ↗ ↻ +	Korpus: 16	16	3	13
ZEIT7C	↘ ↗ ↻ +	Korpus: 129	129	117	12

Was sieht man?

FREIZEIT → klar zweihändig

STUNDE → einhändig

ZEIT → beides, Bild nicht so eindeutig: ZEIT7C eher einhändig, 7A und 7B eher zweihändig

Weitere Fragen an Daten:

z.B. ZEIT7C Tendenz zur Einhändigkeit bedingt durch einhändige Grundform und Annotationsbearbeitungsstand (d.h. abweichende Formen noch nicht aufgelöst)? → Auftrag an Transkription

Lexikographische Entscheidungen:

gleiche oder getrennte Einträge?

Grundform ein- oder zweihändig?

Welche Kommentare?

z.B. kann ein- und zweihändig ausgeführt werden, in der Bedeutung ‚Freizeit‘ (LA ...) jedoch fast immer zweihändig

z.B.: LA ...: Variante mit ausgestrecktem Zeigefinger... meist einhändig

Arbeitsschritte: 2. Sichtung und 3. Lemmakonstitution / Variantensets

Umgebung auf Lexemebene prüfen

- Gibt es Lexeme, die als Varianten noch mit berücksichtigt werden müssen? (Variantenset)
- Gibt es Lexeme, die zu verwandten Gebärden gehören? (→ Verweis)
- Gibt es Lexeme, die (teil)synonym sind? (→ Verweis)

Korpus Lexeme (mit ererbten Tokens) + Parent

Q zeit

43 Einträge

Glosse	^	Tokens	HamNoSys	Parent
\$NUM-JAHR-NACH-JETZT1		34	110 [1 0] 1	\$NUM-ZEIT-NACH-JETZT1
\$NUM-JAHR-VOR-JETZT1		75	110 [1 0] 1	\$NUM-ZEIT-VOR-JETZT1
\$NUM-WOCHE-NACH-JETZT1		14	110 [1 0] 1	\$NUM-ZEIT-NACH-JETZT1
\$NUM-WOCHE-VOR-JETZT1		14	110 [1 0] 1	\$NUM-ZEIT-VOR-JETZT1
FREIZEIT1A		1	110 [1 0] 1	ZEIT7B-\$\$SAM
FREIZEIT1B		21	110 [1 0] 1	ZEIT7A-\$\$SAM
FREIZEIT1C		4	110 [1 0] 1	ZEIT7C-\$\$SAM
HALBZEIT2		1	110 [1 0] 1	UM1A-\$\$SAM
HALBZEIT3		2	110 [1 0] 1	WECHSELN2-\$\$SAM
HOCHZEIT1		17	110 [1 0] 1	SCHLEIER1-\$\$SAM
HOCHZEIT2A		47	110 [1 0] 1	HEIRATEN3A-\$\$SAM
HOCHZEIT2B		2	110 [1 0] 1	HEIRATEN3B-\$\$SAM
HOCHZEIT3		5	110 [1 0] 1	HEIRATEN1-\$\$SAM
KEINE-ZEIT1		3	110 [1 0] 1	ZEIT1-\$\$SAM
LANGE1A		61	110 [1 0] 1	LANG-ZEIT1A-\$\$SAM
LANGE1B		15	110 [1 0] 1	LANG-ZEIT1B-\$\$SAM
LANGE1C		24	110 [1 0] 1	LANG-ZEIT1C-\$\$SAM
LANGE1D		3	110 [1 0] 1	LANG-ZEIT1D-\$\$SAM
LANGE2		21	110 [1 0] 1	LANG-ZEIT2-\$\$SAM
LANGE6		3	110 [1 0] 1	LANG-ZEIT3-\$\$SAM
LESEN-ZEITUNG1		7	110 [1 0] 1	LESEN-ZEITUNG1-\$\$SAM
SOMMER16		4	110 [1 0] 1	JAHRESZEIT1-\$\$SAM
STUNDE1A		35	110 [1 0] 1	ZEIT7C-\$\$SAM
STUNDE2		31	110 [1 0] 1	ZEIT5-\$\$SAM
UHR1		157	110 [1 0] 1	ZEIT1-\$\$SAM
UHR6		1	110 [1 0] 1	ZEIT6-\$\$SAM
WIE-LANGE1A		18	110 [1 0] 1	LANG-ZEIT1A-\$\$SAM
WIE-LANGE1B		11	110 [1 0] 1	LANG-ZEIT1C-\$\$SAM
WIE-LANGE2		5	110 [1 0] 1	LANG-ZEIT2-\$\$SAM
ZEIT1		313	110 [1 0] 1	ZEIT1-\$\$SAM
ZEIT3		14	110 [1 0] 1	ZEIT3-\$\$SAM
ZEIT5		168	110 [1 0] 1	ZEIT5-\$\$SAM
ZEIT6		12	110 [1 0] 1	ZEIT6-\$\$SAM
ZEIT7A		63	110 [1 0] 1	ZEIT7A-\$\$SAM
ZEIT7B		16	110 [1 0] 1	ZEIT7B-\$\$SAM
ZEIT7C		129	110 [1 0] 1	ZEIT7C-\$\$SAM
ZEIT9		12	110 [1 0] 1	ZEIT9-\$\$SAM
ZEITRAUM1		3	110 [1 0] 1	PHASE1E-\$\$SAM
ZEITSCHRIFT1		2	110 [1 0] 1	ZEITUNG1B-\$\$SAM
ZEITSCHRIFT2		5	110 [1 0] 1	AUFKLAPPEN-BUCH1-\$\$SAM
ZEITUNG1A		57	110 [1 0] 1	ZEITUNG1A-\$\$SAM
ZEITUNG1B		23	110 [1 0] 1	ZEITUNG1B-\$\$SAM
ZEITUNG2		3	110 [1 0] 1	AUFKLAPPEN-BUCH1-\$\$SAM

Korpus Lexeme (mit ererbten Tokens) + Parent

Q stunde

5 Einträge

Glosse	^	Tokens	HamNoSys	Parent
HALBE-STUNDE1		14	110 [1 0] 1	ZEIGER1-\$\$SAM
STUNDE1A		35	110 [1 0] 1	ZEIT7C-\$\$SAM
STUNDE1B		276	110 [1 0] 1	STUNDE1B-\$\$SAM
STUNDE2		31	110 [1 0] 1	ZEIT5-\$\$SAM
ÜBERSTUNDEN1		1	110 [1 0] 1	ÜBERSTUNDEN1-\$\$SAM

Sicht: Lexemliste



Arbeitsschritte: 2. Sichtung und 3. Lemmakonstitution / "Wortformen" &

Variantensets

Subtypes (inkl. Wortformen/Varianten)

ZEIT7A/B/C-\$\$SAM

Lexeme:

ZEIT (ein- und zweihändig)

FREIZEIT (zweihändig)

Formen:

umgekehrte Drehrichtung

Phasen variieren etwas

STUNDEIA (einhändig)

nur bei ZEIT7C-\$\$SAM

zahleninkorporierend

umgekehrte Drehrichtung

anderes Variantenset

Kommentar_Eintrag_Produkt

Der Daumen kann angelegt oder abgespreizt sein. Die Drehrichtung kann auch umgekehrt sein. Die Gebärde kann zweihändig und einhändig ausgeführt werden. Form 360.1 wird meist einhändig ausgeführt. Die Gebärde wird außer in der Bedeutung 'Freizeit' vor allem von älteren Gehörlosen verwendet.

Art	level	glosse	hamnosys	corpus	andere	alle
Lexeme wie GF	1	FREIZEIT1C	~ ɔ̣.ə.ɔ̣+	1		1
Lexeme-WF (L wie GF)	0	FREIZEIT1C'movdir:Kreis gg Uhr'phs:1	~ ɔ̣.ə.ɔ̣	2		2
Lexeme-WF (L wie GF)	0	FREIZEIT1C'phs:1	~ ɔ̣.ə.ɔ̣	1		1
Gebärde WF m Lexemen	2	ZEIT7C-\$\$SAM'phs:1	ɔ̣.ə.ɔ̣			0
Lexeme andere GF	1	STUNDE1A	ɔ̣.ə.ɔ̣	11	3	14
Lexeme andere GF WF	0	STUNDE1A'assim	~ ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	STUNDE1A'bas:copy	[ɔ̣.ə.ɔ̣.ɔ̣.ɔ̣](C)		1	1
Lexeme andere GF WF	0	STUNDE1A'movdir:Kreis gg Uhr'phs:2	ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	STUNDE1A'phs:2	ɔ̣.ə.ɔ̣+	4	6	10
Lexeme andere GF WF	0	STUNDE1A'q:2	ɔ̣.ə.ɔ̣	4		4
Lexeme andere GF WF	0	STUNDE1A'q:2'movdir:Kreis gg Uhr	ɔ̣.ə.ɔ̣	1		1
Lexeme andere GF WF	0	STUNDE1A'q:2'movdir:Kreis gg Uhr'phs:2	ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	STUNDE1A'q:2'phs:2	ɔ̣.ə.ɔ̣+	5		5
Lexeme andere GF WF	0	STUNDE1A'q:2d'movdir:Kreis gg Uhr	ɔ̣.ə.ɔ̣	1		1
Lexeme andere GF WF	0	STUNDE1A'q:3d	ɔ̣.ə.ɔ̣	1		1
Lexeme andere GF WF	0	STUNDE1A'q:3d'movdir:Kreis gg Uhr'phs:2	ɔ̣.ə.ɔ̣+		2	2
Lexeme andere GF WF	0	STUNDE1A'q:3d'phs:2	ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	STUNDE1A'q:4	ɔ̣.ə.ɔ̣			0
Lexeme andere GF WF	0	STUNDE1A'q:4'movdir:Kreis gg Uhr	ɔ̣.ə.ɔ̣	1		1
Lexeme andere GF WF	0	STUNDE1A'q:6d	[ɔ̣.ə.ɔ̣.ɔ̣.ɔ̣](C)	1		1
Gebärde WF m Lexemen	2	ZEIT7C-\$\$SAM'hd:1	ɔ̣.ə.ɔ̣+	3	5	8
Lexeme andere GF	1	ZEIT7C	ɔ̣.ə.ɔ̣+	28	23	51
Lexeme andere GF WF	0	ZEIT7C'assim	~ ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	ZEIT7C'hd:2	~ ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	ZEIT7C'hd:2'movdir:Kreis gg Uhr	~ ɔ̣.ə.ɔ̣+		2	2
Lexeme andere GF WF	0	ZEIT7C'hd:2'movdir:Kreis gg Uhr'phs:1	~ ɔ̣.ə.ɔ̣	1	4	5
Lexeme andere GF WF	0	ZEIT7C'hd:2'phs:1	~ ɔ̣.ə.ɔ̣	9	4	13
Lexeme andere GF WF	0	ZEIT7C'hd:2x	~ ɔ̣.ə.ɔ̣+	1		1
Lexeme andere GF WF	0	ZEIT7C'movdir:Kreis gg Uhr	ɔ̣.ə.ɔ̣+	6	3	9
Lexeme andere GF WF	0	ZEIT7C'movdir:Kreis gg Uhr'phs:1	ɔ̣.ə.ɔ̣	26	7	33
Lexeme andere GF WF	0	ZEIT7C'phs:0	ɔ̣.ə.ɔ̣	1		1
Lexeme andere GF WF	0	ZEIT7C'phs:0,5	ɔ̣.ə.ɔ̣	7		7
Lexeme andere GF WF	0	ZEIT7C'phs:1	ɔ̣.ə.ɔ̣	47	25	72
Lexeme andere GF WF	0	ZEIT7C'phs:mehrfach	ɔ̣.ə.ɔ̣+	1	1	2
Gebärde	3	ZEIT7C-\$\$SAM	~ ɔ̣.ə.ɔ̣+			0

Glosse	Tokens	HamNoSys	Parent
STUNDE1B	274	ɔ̣.ə.ɔ̣	STUNDE1B-\$\$SAM
STUNDE1A	35	ɔ̣.ə.ɔ̣	ZEIT7C-\$\$SAM

Art	level	glosse	hamnosys	corpus	andere	alle
Lexeme wie GF	1	ZEIT7A	~ ɔ̣.ə.ɔ̣+	32	101	133
Lexeme-WF (L wie GF)	0	ZEIT7A'assim	~ ɔ̣.ə.ɔ̣+	1		1
Lexeme-WF (L wie GF)	0	ZEIT7A'hd:1	ɔ̣.ə.ɔ̣+	1		1
Lexeme-WF (L wie GF)	0	ZEIT7A'hd:1'phs:1	ɔ̣.ə.ɔ̣	11		11
Lexeme-WF (L wie GF)	0	ZEIT7A'movdir:Kreis gg Uhr'phs:1	~ ɔ̣.ə.ɔ̣		1	1
Lexeme-WF (L wie GF)	0	ZEIT7A'phs:0	ɔ̣.ə.ɔ̣	1		1
Lexeme-WF (L wie GF)	0	ZEIT7A'phs:1	~ ɔ̣.ə.ɔ̣	17	3	20
Lexeme wie GF	1	FREIZEIT1B	~ ɔ̣.ə.ɔ̣+	9	4	13
Lexeme-WF (L wie GF)	0	FREIZEIT1B'movdir:Kreis gg Uhr'phs:1	~ ɔ̣.ə.ɔ̣		3	3
Lexeme-WF (L wie GF)	0	FREIZEIT1B'phs:1	~ ɔ̣.ə.ɔ̣	12	8	20
Gebärde	3	ZEIT7A-\$\$SAM	~ ɔ̣.ə.ɔ̣+	1	1	2

lexikographische Entscheidung (Lemmakonstitution):

→ eigener Eintrag für STUNDE (Form 1B häufiger als 1A)

361

- Varianten**
- 1 ilex://types.id=54108
STUNDE1B-\$\$SAM
ɔ̣.ə.ɔ̣
 - 2 ilex://types.id=54106
STUNDE1A
ɔ̣.ə.ɔ̣

Kommentar_Eintrag_Produkt

Kreisbewegung kann auch in entgegengesetzte Richtung gehen. Bei inkorporierter 1 auch die Handform mit dem Daumen.

Grammatik

Zahleninkorporierende Gebärde

Art	level	glosse	hamnosys	corpus	andere	alle
Lexeme wie GF	1	ZEIT7B	~ ɔ̣.ə.ɔ̣+	10	23	33
Lexeme-WF (L wie GF)	0	ZEIT7B'hd:1'phs:1	ɔ̣.ə.ɔ̣	2		2
Lexeme-WF (L wie GF)	0	ZEIT7B'phs:1	~ ɔ̣.ə.ɔ̣	3	2	5
Lexeme wie GF	1	FREIZEIT1A	~ ɔ̣.ə.ɔ̣+	1	5	6
Lexeme-WF (L wie GF)	0	FREIZEIT1A'phs:1	~ ɔ̣.ə.ɔ̣	1	6	7
Lexeme-WF (L wie GF)	0	FREIZEIT1A'phs:2	~ ɔ̣.ə.ɔ̣+			0
Gebärde	3	ZEIT7B-\$\$SAM	~ ɔ̣.ə.ɔ̣+		1	1
Gebärde WF o Lexeme	2	ZEIT7B-\$\$SAM'phs:1	~ ɔ̣.ə.ɔ̣	1		1

Arbeitsschritte: 3. Lemmakonstitution und 6. Lesartendifferenzierung

Ergebnis: 2 Einträge

360

Varianten	1	2	3	4	5
	ilex: ZEIT7C-SSAM	ilex: ZEIT7A-SSAM	ilex: ZEIT7B-SSAM		
<p>1 Stunde wurde ausgelagert (siehe 361), da Zahleninkorporation eher einhändig, keine Varianten...</p> <p>Freizeit: zahlenmäßig eher die L hand</p> <p>Zeit: zahlenmäßig eher die Zeigefingerhand</p> <p>Notiz zu besprechen</p> <p>1 Runde</p> <p>Formgleich 1: 361 -- nur eine Variante-- 361 ist in iLex teil on ZEIT7C-SSAM, bei uns ausgelagert -- wie hier angeben?</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>Red Sitzung: hauptvariante (einhändig/zweihändig problem)</p>					
<p>wartet auf Studioauf</p> <p>fast fertig</p> <p>akut in Arbeit</p> <p>in Arbeit</p> <p>reserviert</p> <p>nach Glossenprüfen nochmal</p> <p>Fragen für Red-Sitzung</p> <p>wartet auf mehr Korpusstokens</p> <p>Kurzeintrag vor abgeschlossen</p> <p>bekommt Vollentrag</p> <p>nach Produktion anschauen</p>					
<p>ck Filme Tagassets pro Eintrag</p> <p>Export Paket an Thomas Vorher: Aktualisierung der Gebärdenliste</p> <p>1. Suche: Produzieren ja</p> <p>2. globalfelder leeren</p> <p>3a. ilex-Teile für aufgewasene Einträge vorbereiten</p> <p>3b. mit fgl zuordnen</p> <p>3c. mit fgl + Knoten</p> <p>Bei allen Paketzusammenstellungen alle anderen raus aus FM</p> <p>ck Filme Tagassets ohne stills für Paket (alle aufgerufenen DS) öffnet tag</p> <p>ck Filme Tagassets für Paket (alle aufgerufenen DS) öffnet tag</p> <p>1. suche: Prod: ja</p> <p>2. Einträge vorbereiten (Bspids)</p> <p>3b. Paket vorb + assets</p> <p>4 global -> in iLex</p> <p>4. global -> in iLex</p>					
<p>alle Belege des eintrags iLEX bearb</p> <p>alle Belege des eintrags iLEX Ergebnis</p> <p>produktions-relevante Bel.</p> <p>produktions-relevante Bel.</p> <p>alle Belege + Details (langsam)</p> <p>wartet auf</p> <p>5. ->Export als html speichern</p> <p>x anschauen ob alle Einträge auf ersten Blick plausibel</p>					
<p>LA Def im Prod</p> <p>F UA im Produkt</p> <p>Bsp gewählt fertig bearb</p> <p>LA Notiz zu besprechen</p> <p>Lumping</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. ZEIT1</p> <p>Syn2 ilex:/types. PHASE1A</p> <p>Syn3 ilex:/types. PHASE2B</p> <p>Syn4</p> <p>Syn5</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2364</p> <p>Lebensabschnitt oder Lebensphase, die als</p> <p>ein meist längerer Zeit- oder Lebensabschnitt, der als abgeschlossen empfunden wird und durch einen bestimmten Aspekt charakterisiert werden kann, z.B. Lebensabschnitt, Ausbildungsabschnitt, historische Epoche; oft als Teil einer aus</p> <p>Zeit</p> <p>...zeit</p> <p>Lebensabschnitt</p> <p>Epoche</p> <p>MB_P zeit</p> <p>Zeit</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>2317 4 Beispiele um unterLA und Varianten abzudecken!</p> <p>WEIHNACHTSZEIT Beleg 2</p> <p>SCHULZEIT Beleg 1 - Beleg 6</p> <p>DDR-ZEIT (Atom) Beleg 3</p> <p>MITTEL ALTER ZEIT Beleg 7</p> <p>SCHWERE ZEIT Beleg 5</p> <p>STEIN ZEIT Beleg 8</p> <p>3 Beispiele, weil 1 Bsp mit FREI und ZEIT dabei</p> <p>longform short form Überlegung ob gemeinsames Muster mit FREI1 ZEIT1</p> <p>3 Beispiele</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. JETZT1</p> <p>Syn2 ilex:/types. JETZT3</p> <p>Syn3 ilex:/types. reg JETZT2</p> <p>Syn4 ilex:/types. HEUTE</p> <p>Syn5 ilex:/types. HEUTE3</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2365</p> <p>Zeit, die einem außerhalb der beruflichen Tätigkeit zur freien Verfügung steht</p> <p>Zeit, die einem außerhalb der beruflichen Tätigkeit zur freien Verfügung steht</p> <p>Freizeit</p> <p>MB_P freizeit</p> <p>Zeit</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>3 Beispiele, weil 1 Bsp mit FREI und ZEIT dabei</p> <p>longform short form Überlegung ob gemeinsames Muster mit FREI1 ZEIT1</p> <p>3 Beispiele</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. gl ZEIT1</p> <p>Syn2</p> <p>Syn3</p> <p>Syn4</p> <p>Syn5</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2366</p> <p>Zeitraum oder -dauer, die für etwas zur Verfügung stehen oder</p> <p>Zeitraum oder -dauer, die für etwas zur Verfügung stehen oder benötigt wird</p> <p>Zeit</p> <p>MB_P zeit</p> <p>Zeit</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>3 Beispiele</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. JETZT1</p> <p>Syn2 ilex:/types. JETZT3</p> <p>Syn3 ilex:/types. reg JETZT2</p> <p>Syn4 ilex:/types. HEUTE</p> <p>Syn5 ilex:/types. HEUTE3</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2367</p> <p>Bezugnahme auf den Zeitabschnitt oder Zeitpunkt der unmittelbaren Gegenwart</p> <p>zurzeit im Moment heutzutage heute in der Gegenwart jetzt gerade</p> <p>MB_P zeit</p> <p>Zeit</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>3 Beispiele</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. FRUHER1</p> <p>Syn2 ilex:/types. FRUHER2</p> <p>Syn3 ilex:/types. GESCHICHTE1B</p> <p>Syn4 ilex:/types. GESCHICHTE2</p> <p>Syn5</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2376</p> <p>in der Vergangenheit / früher</p> <p>Bezugnahme auf einen Zeitabschnitt oder Zeitpunkt der meist schon weiter zurückliegenden Vergangenheit</p> <p>früher in der Vergangenheit damals historisch</p> <p>MB_P zeit</p> <p>Zeit</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>3 Beispiele</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. BIS1</p> <p>Syn2</p> <p>Syn3</p> <p>Syn4</p> <p>Syn5</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2377</p> <p>zeigt an, dass zwei (genannte oder implizierte) Ereignisse zeitlich als zu der Zeit</p> <p>als zu der Zeit</p> <p>MB_P zeit</p> <p>Zeit</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>3 Beispiele</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. BIS1</p> <p>Syn2</p> <p>Syn3</p> <p>Syn4</p> <p>Syn5</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					

361

Varianten	1	2	3	4	5
	ilex: STUNDE2B-SSAM	ilex: STUNDE2A-SSAM			
<p>1 Stunde</p> <p>Notiz zu besprechen</p> <p>1 Runde</p> <p>mit Thomas: Formgleich 1: 360</p> <p>Redsitzung: Grundform mit Daumen statt Zeigefinger (3:1 bei Tokens; 2:1 bei Informanten) ? - aber Stunde neutral eher Zeigefinger? (letzteres nicht überprüft)</p> <p>Redsitzung: Formulierung des Grammatikkommentars: Detail? ...</p> <p>wartet auf Studioauf</p> <p>fast fertig</p> <p>akut in Arbeit</p> <p>in Arbeit</p> <p>reserviert</p> <p>nach Glossenprüfen nochmal</p> <p>Fragen für Red-Sitzung</p> <p>wartet auf mehr Korpusstokens</p> <p>Kurzeintrag vor abgeschlossen</p> <p>bekommt Vollentrag</p> <p>nach Produktion anschauen</p> <p>ck Filme Tagassets pro Eintrag</p> <p>Export Paket an Thomas Vorher: Aktualisierung der Gebärdenliste</p> <p>1. Suche: Produzieren ja</p> <p>2. globalfelder leeren</p> <p>3a. ilex-Teile für aufgewasene Einträge vorbereiten</p> <p>3b. mit fgl zuordnen</p> <p>3c. mit fgl + Knoten</p> <p>Bei allen Paketzusammenstellungen alle anderen raus aus FM</p> <p>ck Filme Tagassets ohne stills für Paket (alle aufgerufenen DS) öffnet tag</p> <p>ck Filme Tagassets für Paket (alle aufgerufenen DS) öffnet tag</p> <p>1. suche: Prod: ja</p> <p>2. Einträge vorbereiten (Bspids)</p> <p>3b. Paket vorb + assets</p> <p>4 global -> in iLex</p> <p>4. global -> in iLex</p> <p>alle Belege des eintrags iLEX bearb</p> <p>alle Belege des eintrags iLEX Ergebnis</p> <p>produktions-relevante Bel.</p> <p>produktions-relevante Bel.</p> <p>alle Belege + Details (langsam)</p> <p>wartet auf</p> <p>5. ->Export als html speichern</p> <p>x anschauen ob alle Einträge auf ersten Blick plausibel</p>					
<p>LA Def im Prod</p> <p>F UA im Produkt</p> <p>Bsp gewählt fertig bearb</p> <p>LA Notiz zu besprechen</p> <p>Lumping</p> <p>LA InterneNotizen_Bearbeiterin</p> <p>Syn1 ilex:/types. BIS1</p> <p>Syn2</p> <p>Syn3</p> <p>Syn4</p> <p>Syn5</p> <p>Ant1</p> <p>Ant2</p> <p>Ant3</p> <p>LA Kommentar im Produkt regional</p> <p>LA Tiele fein mittel grob</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2383</p> <p>Zeiteinheit von 60 Minuten</p> <p>Maßeinheit für Zeit (60 Minuten)</p> <p>Stunde</p> <p>MB_P stunde</p> <p>Stunde</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>schönes Beispiel für 2 versch handformen bei inkorp und reiner Maßeinheit ilex:/tags.id=2146795</p> <p>ohne Zahleninkorporation, meist in Verbindung mit einer vorausgehenden Zahlgebärde</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2362</p> <p>zur Angabe einer Zeitdauer genutzt mit inkorporierter</p> <p>genaue Angabe der Dauer oder zeitlichen Länge z.B. einer Aktivität oder einer Veranstaltung</p> <p>... Stunde(n)</p> <p>... Stunde(n) lang</p> <p>in ... Stunden</p> <p>MB_P [zahl] stunde*</p> <p>Stunde</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>inklusive - in ... Stunden</p> <p>- ... Stunden lang</p> <p>3 Beispiele weil verschiedene Formen</p> <p>schönes Beispiel für 2 versch handformen bei inkorp und reiner Maßeinheit ilex:/tags.id=2146795</p> <p>mit Angabe der Stundenzahl über die Handform und Zahl der Hände (Zahleninkorporation)</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					
<p>2627</p> <p>eine ungefähre Zeitdauer von mehreren Stunden, die durch zwei oder mehrere hintereinander in die bewegung integrierte Zahlhandformen etwas genauer eingegrenzt wird</p> <p>... bis ... Stunden</p> <p>innerhalb von ... bis ... Stunden</p> <p>MB_P [zahlA] [zahlB] stunde*</p> <p>Stunden</p> <p>zu besprechen</p> <p>Red Sitzung: UÄ...zeit zu diskutieren</p> <p>eine ungefähre Zeitdauer kann durch das Ausführen der möglichen Stundenzahlen nacheinander</p> <p>LA typische Muster</p> <p>LA typische Muster</p> <p>LA Sem_Synt</p> <p>DWS_status_LA gesichert (ins Produkt)</p>					

Arbeitsschritt: 4. Verweise auf Eintragungsebene

Sicht: Formgleich/Formähnlich-Verweise

Gebärden: ZEIT7C-\$\$SAM							
Form	Kinder	QxQ	Tokens	~	∞		
5 Einträge							
Gebärde	überlapp	Beziehung	Motivation	HamNoSys	analyser	eingetragen	
\$NUM-TEEN5-\$\$SAM	^	überlapp	ungeprüft	∞	OJ	2016-09-13	
JAH1A-\$\$SAM	4	formähnlich	gemeinsame Bildmerkmale	∞	Hk	2004-12-20	
JAH3A-\$\$SAM	3	formähnlich	ungeprüft	∞	SWO	2008-09-30	
SEKUNDE1A-\$\$SAM	4	formähnlich	ähnliche Bilder	∞	Hk	2004-12-20	
TEMPERATUR4-\$\$SAM	1	formähnlich	ähnliche Bilder	∞	IH	2008-09-24	

Gebärden: ZEIT7A-\$\$SAM							
Form	Kinder	QxQ	Tokens	~	∞		
3 Einträge							
Gebärde	überlapp	Beziehung	Motivation	HamNoSys	analyser	eingetragen	
\$NEU-KINDCHENSHEMA1-\$\$SAM	^	formähnlich	verschiedene Bilder	∞	Hk	2004-12-20	
\$NUM-ZWEI1A-\$\$SAM		formähnlich	ungeprüft	∞	AD	2010-03-24	
FEIERN1-\$\$SAM	7	formähnlich	ungeprüft	∞	OJ	2016-12-08	

Gebärden: ZEIT7B-\$\$SAM							
Form	Kinder	QxQ	Tokens	~	∞		
6 Einträge							
Gebärde	überlapp	Beziehung	Motivation	HamNoSys	analyser	eingetragen	
ARBEITEN4-\$\$SAM	8	formähnlich	ungeprüft	∞	GL	2017-01-23	
FEIERN1-\$\$SAM	7	formähnlich	ungeprüft	∞	RK	2016-11-23	
FÖRDERN1A-\$\$SAM	4	formähnlich	ungeprüft	∞	GL	2010-04-17	
GESELLSCHAFT1-\$\$SAM	4	formähnlich	Bildbeziehung unklar	∞	Hk	2004-12-20	
MISCHEN4-\$\$SAM	8	formähnlich	verschiedene Bilder	∞	GL	2010-04-17	
OPFER4-\$\$SAM	2	formähnlich	verschiedene Bilder	∞	Hk	2004-12-20	

formgleich:
Verweis oder in gleichem Eintrag?

formähnlich:
Verweis?

verwandt:
Verweis?

360

Varianten

- 1 ilex://types.id=13746
ZEIT7C-\$\$SAM
∞
- 2 ilex://types.id=14984
ZEIT7A-\$\$SAM
∞
- 3 ilex://types.id=14983
ZEIT7B-\$\$SAM
∞

formgleich1			
formgleich2			
formaehnlich1	ilex://types.	STUNDE1B-\$\$SAM	∞
formaehnlich2	ilex://types.	GESELLSCHAFT1-\$\$SAM	∞
formaehnlich3	ilex://types.	FEIERN1-\$\$SAM	∞
formaehnlich4			
formaehnlich5			
verwandt1			
verwandt2			
verwandt3			

361

Varianten

- 1 ilex://types.id=54108
STUNDE1B-\$\$SAM
∞
- 2 ilex://types.id=54106
STUNDE1A
∞

formgleich1			
formgleich2			
formaehnlich1	ilex://types.	JAH1A-\$\$SAM	∞
formaehnlich2	ilex://types.	ZEIT7C-\$\$SAM	∞
formaehnlich3			
formaehnlich4			
formaehnlich5			
verwandt1	ilex://types.	ZEIT5-\$\$SAM	∞
verwandt2	ilex://types.	SEKUNDE1A-\$\$SAM	∞
verwandt3			

Arbeitsschritt 5: Überblick über Kollokationen und Zusammensetzungen

Kandidaten für Kollokatoren und Zusammensetzungen Sicht: Häufige rechte und linke Nachbarn

Gebärden: ZEIT1-\$SAM

39 Einträge

Position	Kollokator	Basis	Kollokator	Mi-Wert	gem Auftreten	dist Inform gem vk	Kollokator i Korpus	Basis i Korpus	Kollokatoren glossen
links	\$NUM-TEEN	UHR1		6.75	154	59	2469	350	\$NUM-TEEN1 \$NUM-TEEN2A \$NUM-TEEN2B \$NUM-TEEN3 \$NUM-TEEN4B \$NUM-TEEN5 \$NUM-TEEN6A
links	\$NUM-SCHNIPS	UHR1		6.49	10	8	192	350	\$NUM-SCHNIPS1
links	WIE-VIEL	UHR1		6.10	8	7	201	350	WIE-VIEL1 WIE-VIEL5
links	\$NUM-ZEHNER	UHR1		4.61	29	21	2043	350	\$NUM-ZEHNER1 \$NUM-ZEHNER2A \$NUM-ZEHNER2B
links	\$NUM-UHR	UHR1		2.52	6	5	1809	350	\$NUM-UHR1A \$NUM-UHR1B \$NUM-UHR1D
links	\$NUM-EINER	UHR1		1.83	18	13	8753	350	\$NUM-EINER1A \$NUM-EINER1B \$NUM-EINER1D
rechts		UHR1	ABEND	4.13	9	7	883	350	ABEND1 ABEND2
rechts		UHR1	\$NUM-TEEN	4.07	24	21	2469	350	\$NUM-TEEN1 \$NUM-TEEN2A \$NUM-TEEN2B \$NUM-TEEN6A
rechts		UHR1	BIS	4.06	13	8	1340	350	BIS1
rechts		UHR1	\$NUM-UHR	3.63	13	10	1809	350	\$NUM-UHR1A \$NUM-UHR1B \$NUM-UHR1D
rechts		UHR1	FREI	3.52	5	4	753	350	FREI1 FREI4
rechts		UHR1	\$NUM-EINER	2.24	24	19	8753	350	\$NUM-EINER1A \$NUM-EINER1B \$NUM-EINER1C \$NUM-EINER1D
rechts		UHR1	ARBEITEN	2.15	5	4	1939	350	ARBEITEN1 ARBEITEN2
rechts		UHR1	DU	2.05	12	7	5002	350	DU1 DU1-\$SAM
links	TEIL	ZEIT1		5.71	5	3	110	522	TEIL1A TEIL1B
links	BESTIMMT	ZEIT1		5.27	5	5	150	522	BESTIMMT1
links	GLEICH	ZEIT1		4.84	18	14	728	522	GLEICH1A GLEICH1C GLEICH8
links	BRAUCHEN	ZEIT1		4.24	11	10	671	522	BRAUCHEN1
links	KEIN	ZEIT1		4.03	13	8	920	522	KEIN1 KEIN2 KEIN3
links	SCHÖN	ZEIT1		3.63	15	11	1404	522	SCHÖN1A SCHÖN1B SCHÖN3 SCHÖN5
links	SCHNELL	ZEIT1		3.33	5	5	573	522	SCHNELL2 SCHNELL3A SCHNELL3B SCHNELL5 SCHNELL6
links	MEHR	ZEIT1		3.18	13	11	1661	522	MEHR1 MEHR3 MEHR5
links	FREI	ZEIT1		2.94	5	5	753	522	FREI1 FREI2A
links	JAHR	ZEIT1		2.32	6	6	1384	522	JAHR1A JAHR1B JAHR2A JAHR4A JAHR4B
links	VIEL	ZEIT1		2.27	9	8	2155	522	VIEL1A VIEL1B VIEL5
links	BIS	ZEIT1		2.11	5	4	1340	522	BIS1 BIS1-\$SAM
links	ARBEITEN	ZEIT1		2.06	7	7	1939	522	ARBEITEN1 ARBEITEN2 ARBEITEN4
links	GUT	ZEIT1		1.46	8	7	3358	522	GUT1 GUT1-\$SAM GUT3
links	WIE	ZEIT1		1.39	6	5	2644	522	WIE3B WIE4A
links	FERTIG	ZEIT1		1.23	5	5	2459	522	FERTIG1A FERTIG1B FERTIG2
links	MEIN	ZEIT1		1.19	5	4	2528	522	MEIN1
rechts		ZEIT1	KNAPP	7.39	8	6	55	522	KNAPP1
rechts		ZEIT1	DRUCK-AUF-PERSON	6.14	6	5	98	522	DRUCK-AUF-PERSON1
rechts		ZEIT1	ENTWICKELN	4.27	5	4	299	522	ENTWICKELN1A ENTWICKELN1B ENTWICKELN2
rechts		ZEIT1	SCHNELL	3.33	5	5	573	522	SCHNELL1A SCHNELL1B SCHNELL2 SCHNELL3A SCHNELL6
rechts		ZEIT1	EGAL	3.32	5	4	579	522	EGAL3
rechts		ZEIT1	FÜR	2.82	10	8	1641	522	FÜR1
rechts		ZEIT1	\$NUM-UHR	1.94	6	5	1809	522	\$NUM-UHR1A \$NUM-UHR1B \$NUM-UHR1D
rechts		ZEIT1	MUSS	1.25	5	5	2427	522	MUSS1

beim Eintrag (Zusammensetzungen):

Zusammensetzungen u Wendungen, Komposita FB

TEIL1A ZEIT1 (Teillzeit)
 [ARBEITEN1 / ARBEITEN2] ZEIT1 (Arbeitszeit)
 [FREI1 / FREI2A] ZEIT1 (Freizeit)
 ZEIT1 DRUCK-AUF-PERSON1 (Zeitdruck)

bei der Lesart (Kollokationen):

LA_typische_Muster -->Muster Be

[Zahlgebärde] UHR1
 [WIE-VIEL1 / WIE-VIEL5] UHR1



Arbeitsschritt 5: Überblick über Kollokationen und Zusammensetzungen

Zusammensetzungen / Kollokationsmuster / Phrasen

Layout: Muster u...earbeiten | Anzeige: [Grid] [List] [Table] | Seitenansicht | Helvetica | Normal | 16 pt

M_bearb: GL, AM, SaW, RK

M_Regieliste: aufnehmen, aufgenommen, angebunden (Phrase), soll nicht auf Regieliste, in Arbeit, noch leer

M_Korpus_dictproduktion: nein

F_gesamtcheck

M_id: 19

M_deutsch: Bei Lehnkomposita und Run-ons mit analoger Struktur zum Deutschen zB, Arbeitszeit

M_kommentar_intern

M_nur_Text_ins_Produkt ohne Glossen/IDs: Bei Lehnkomposita und Run-ons mit analoger Struktur zum Deutschen

M_A_Beschreibung: charakterisierendes Element

M_B_Beschreibung

M_signA...	MB für Studio M_a..._mb	F ck a... _film_vhd	M_A..._rep_tag	M_signB...	MB für Studio M_b..._mb	F ck b..._film_vhd	M_B..._rep_tag
A1 illex: SCHULE1A	schule	Film vhd	1 [Schule]	B1 illex: ZEIT7A	zeit	Film vhd	1 [Zeit]
A2 illex: KRIEG1A	krieg	Film vhd	2 [Krieg]	B2			2
A3 illex: DDR1A	ddr	Film vhd	3 [DDR]	B3			3
A4 illex: DDR2	ddr	Film vhd	4 [DDR]	B4			4
A5 illex: DDR3	ddr	Film vhd	5 [DDR]	B5			5
A6			6	B6			6
A7			7	B7			7
A8			8	B8			8
A9			9	B9			9
A10			10	B10			10

bidirektionaler Pfeil ja

F_M_Muster_fuer_layout: [charakterisierendes Element: SCHULE1A / KRIEG1A / DDR1A / DDR2 / DDR3] ZEIT7A

F_M_Muster_fuer_thomas_export: [charakterisierendes Element: 6130,9578(bag=416),36951,55439,87116] 7451(bag=360)

M_Konzept_ID: iLex Konzept für die Phrase im Transkript

M_filmtag: [-> film (illex) öffnen]

Verlinkt als idiomsche Phrase bei Folgenden LA:

Verlinkt als Muster bei Folgenden LA: 360 ZEIT7C-\$SAM LA: 2364 e Zeit- oder Lebensabschnitt, der empfunden wird und durch einer charakterisiert werden kann, z.B. Ausbildungsabschnitt, historisch einer aus mehreren Gebärdens zu Benennung dieses Abschnitts oc

Darstellung im Produkt

HÄUFIGE KOMBINATIONEN

06130 416 036951 055439 087116

charakterisierendes Element

7451



Arbeitsschritt 6b: Sichten und Analyse einzelner Belege

Sicht: Token-Tab beim Type mit für Arbeitsschritt relevanten Angaben

LA-Nr Lesartennotiz weitere Notizen Übersetzung

Form	Bedeut.	Kinder	Tokens	~	∞	∞	Sprache	Standb.
Lexeme: SAUBER1								
188 Einträge								
Hand	Glosse	Inform...	Substask	Teilkorpus	workflow	LA in...	Hamnosys	red Bsp
r	SAUBER1	BER-12	Freie Konversati...	AB veröffentli...		1104		
r	SAUBER1	KOE-05	Ausgewählte Ve...	A veröffentlic...				
r	SAUBER1	STU-14	Unterhaltung	AB veröffentli...				
r	SAUBER1	HB-14	9. Schild					
l	SAUBER1	HH-16	Terminabsprache					
r	SAUBER1	NUE-43	Diskussion	AB veröffentli...		2541		
r	SAUBER1	BER-28	Behörden	AB veröffentli...		2541		
r	SAUBER1	MST-19	Religion	AB veröffentli...		1107		
r	SAUBER1	KOE-05	Ausgewählte Ve...	A veröffentlic...				
r	SAUBER1	MUE-29	GARTEN_item					
r	SAUBER1	MUE-25	10. Schild					
r	SAUBER1	FRA-36	Freie Konversati...					
r	SAUBER1	MST-40	Diskussion	AB				
r	SAUBER1	MUE-50	Gehörlosenthe...	AB veröffentli...		1108 ...		
r	SAUBER1	BER-17	Stadt und Land		alles fertig	1101		Thema: Vergleich zw...
r	SAUBER1	FRA-36	Verkehr					
r	SAUBER1	BER-53	Gehörlosenthe...					
r	SAUBER1	FRA-36	Ausgewähltes T...					
r	SAUBER1	GOE-20	Politik	AB veröffentli...		1107 ...		
r	SAUBER1	HH-05	SCHMUTZIG_item					
r	SAUBER1	MVP-19	Freie Konversati...			1101		
r	SAUBER1	NUE-37	Unterhaltung	AB veröffentli...				
r	SAUBER1	MUE-40	Freie Konversati...	AB veröffentli...				
r	SAUBER1	HH-33	Sport und Spiel			1109?		
l	SAUBER1	LEI-56	Ausgewähltes T...	AB veröffentli...		1104		
l	SAUBER1	HH-25	SCHMUTZIG_item					
r	SAUBER1	KOE-31	Ausgewähltes T...	AB veröffentli...	alles fertig	1104		An der Wand konnte...
r	SAUBER1	HH-24	Körperpflege, G...					
r	SAUBER1	HH-24	Körperpflege, G...					
r	SAUBER1	MST-19	Religion	AB veröffentli...	alles fertig	1107		
r	SAUBER1	GOE-36	Ausgewählte Ve...	AB veröffentli...	alles fertig	2540		Nach einem Arztgesp...
r	SAUBER1	MST-107	SCHMUTZIG_item					
r	SAUBER1	BER-30	Ausgewähltes T...			2540		
r	SAUBER1	BER-35	Diskussion	AB				
r	SAUBER1	KOE-20	Freie Konversati...	AB veröffentli...	Beispiel verw...	2539		Thema: Ein selbst au...
r	SAUBER1	BER-13	Freie Konversati...	AB veröffentli...				
l	SAUBER1	HH-25	Arbeit und Beruf					
r	SAUBER1	FRA-43	Ausgewählte Ve...	AB veröffentli...		1104		
r	SAUBER1	HB-14	9. Schild					
r	SAUBER1	KOE-21	Familie und Ver...					
r	SAUBER1	LEI-06	Gehörlosenthe...	AB veröffentli...				
r	SAUBER1	MUE-14	Terminabsprache					
r	SAUBER1	FRA-55	Verkehr		alles fertig	1109		Thema: Vergleich des...
r	SAUBER1	GOE-33	Diskussion	AB veröffentli...		1109		
l	SAUBER1	HH-25	Arbeit und Beruf					
r	SAUBER1	MUE-01	WASSER_item					
r	SAUBER1	MUE-50	14. Schild					
r	SAUBER1	MVP-05	DDR	AB veröffentli...		1104		
r	SAUBER1	BER-53	Gehörlosenthe...					
r	SAUBER1	BER-55	Gehörlosenthe...					
r	SAUBER1	HB-01	10. Schild			1104		
r	SAUBER1	HH-31	Arbeit und Beruf					
r	SAUBER1	BER-36	Gehörlosenthe...	AB veröffentli...	alles fertig	2539		Thema: Vorbilder. De...
r	SAUBER1	KOE-53	Körperpflege, G...	AB veröffentli...		1101		
r	SAUBER1	MST-107	SCHUH_item					
r	SAUBER1	STU-22	Diskussion	AB veröffentli...				
r	SAUBER1	GOE-20	Ausgewählte Ve...			1109?		
r	SAUBER1	MUE-58	10. Schild					
r	SAUBER1	MUE-58	10. Schild					
r	SAUBER1	HB-14	Ausgewähltes T...	AB veröffentli...				
r	SAUBER1	BER-53	Gehörlosenthe...					
r	SAUBER1	HH-24	Körperpflege, G...					
r	SAUBER1	MST-32	Gehörlosenthe...					
r	SAUBER1	FI-32	Politik			1109		
1999-06-30 (Susanne König), 2016-02-17 (Julian Bleicken)								

Analyse einzelner Tokens

(Sichten: Transkript mit Film)

Timecode	Segm...	Deut...	Engli...	Kom...	Deut...	Engli...	Kom...	Moderator	Engli...	Segm...	Lex...	Ham...	Ham...	Mund...	Beispiel_B
10:55:35:47 10:55:36:06										X	ORAL1'ph		ʔrə	oral	
10:55:36:06 10:55:36:14										X	NEIN3B-\$		ˈnɪn		
10:55:36:14 10:55:36:23															
10:55:36:23 10:55:36:24															
10:55:36:24 10:55:36:38					Es war eine harte Zeit.	It was a hard time.				X	FRÜHER1		fr̩ɛhɐ		
10:55:36:38 10:55:36:41															
10:55:36:41 10:55:37:04															
10:55:37:04 10:55:37:18										X	PHASE1D	ˈfɑːzə	ˈfɑːzə	bis	
10:55:37:18 10:55:37:24															
10:55:37:24 10:55:37:40										X	ZEIT7C'h	ˈtʰaɪt	ˈtʰaɪt	zeit	
10:55:37:40 10:55:37:49															
10:55:37:49 10:55:38:10										X	HART1A'h		ˈhɑːt	hart	
10:55:38:10 10:55:38:15															
10:55:38:15 10:55:38:29										X	ICH1		ɪç		
10:55:38:29 10:55:38:34															
10:55:38:34 10:55:38:44										X	PHASE1D	ˈfɑːzə	ˈfɑːzə		
10:55:38:44 10:55:39:04															
10:55:39:04 10:55:40:05					Als ich fünfzehn war, wechselte ich zur Berufsschule.	I switched to vocational school by age 15.				X	UND-DAN		ˈdʌn	dann	
10:55:40:05 10:55:40:12															
10:55:40:12 10:55:40:40															
10:55:40:40 10:55:41:03										X	BIS1		ˈbɪs	bis	
10:55:41:03 10:55:41:10															
10:55:41:10 10:55:41:19										X	ICH1	ɪç	ɪç		
10:55:41:19 10:55:41:26															
10:55:41:26 10:55:41:40										X	\$NUM-TE		fʏnfzɛhn	fünfzehn	
10:55:41:40 10:55:42:00															
10:55:42:00 10:55:42:02										X	ALT1A'ph		ˈalt	alt	
10:55:42:02 10:55:43:19															



Analyse einzelner Tokens (Sicht: Tokens in Kontext, Film)

Tokens im Kontext: ZEIT7C

	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
	GEHÖRLOS1A	IDENTITÄT1A	BEDEUTUNG1'phs:1	KONFLIKT1C	HÖREND1A	WELT1	VIEL6	ZEIGEN1A'phs:2	SCHÖN1A	ZEIT7C	MEIN1	LEBEN1E	BIS-JETZT3B	\$GEST-OFF-\$SAM	\$ORAL-\$SAM	MEIN1'phs:2
	UND2A'assim	STAFFEL4	\$NUM-MAL1'q:2	STAFFEL4	\$NUM-MAL1'q:2	BRONZE-MEDAILLE1	HOLEN1A	\$NUM-MAL1'q:2	\$INDEX1'hd:2'loc	ZEIT7C'hd:2'phs:1	VOR1G	\$INDEX1'phs:0	HART1A'bas:copy	\$SPEZIAL-SEHR1-\$SAM	ERDBEBEN1	\$INDEX1'phs:0
	\$GEST-NM-SPRECHEN	KOMMUNIKATION1A'hd	ICH2	\$GEST-KOPFNICKEN1-	SCHREIBEN1A	ORAL1'phs:1	NEIN3B-\$SAM'hd:2	FRÜHER1'phs:1	PHASE1D-\$SAM	ZEIT7C'hd:2'phs:1	HART1A'bas:copy	ICH1	PHASE1D-\$SAM	UND-DANN2	BIS1	ICH1
	FÜHLEN1A	VATER1'movdir:ba_rev	FÜHLEN1A'h_s	\$GEST-OFF-\$SAM'h_s	UNTERSCHIED1C	\$GEST-OFF-\$SAM	RECHT1'hd:1	\$GEST-OFF-\$SAM'hd:1	SCHWER1B'hd:1	C'movdir:Kreis_gg_Uhr'	\$INDEX1'loc_ts_h:links					
	MÄDCHEN1'phs:1	AUFTEILEN3	ORT1A-\$SAM'hd:2	ORT1A-\$SAM'hd:2	\$INDEX1	SCHÖN1A	SCHULE1C'phs:0	TOLL1'phs:1	SCHULE1C'phs:0	ZEIT7C'phs:0,5	JA1A	BEIDE1A'1incl'phs:1	\$GEST-\$SAM'hd:1	ORT1A-\$SAM'hd:2alt	BEIDE1A'1incl'phs:1	WIE2'hd:1
Subtask											Gehörlosenthema 1					
Deutsche Übersetz...	Ich habe mich dem lautsprachlichen Unterricht angepasst. Ich schrieb und sprach.										Es war eine harte Zeit.					
Mundbild/Mundgest...	[MG]	[MG]		schreiben	oral			bis	zeit	hart		dann	bis			
Thema_Schlagworte	Berlin:SuL,Beruf:AuB,Berufsschule:AuB,Ertauben:KGK/LA,Familie:FuV,Internat:SuU,Jugend:LA,Kinder:FuV,Kindergarten:LA/AuB,Orale Erziehung:SuU,Schulzeit:SuU/LA,Sport:SuS,Sportverein:SuS															
Projektkontext	DGS-Korpus : dgskorpus_ber_04 : BER-15 : 2 : dgskorpus_ber_04															



Arbeitsschritt 6: Lesarten-Differenzierung (Word Sense Discrimination WSD)

Definieren und Beschreiben der Lesarten in FileMaker, Zwischenergebnisse festhalten

The screenshot displays a FileMaker database interface for 'FMLEsarten2020-03-03 (SLM-IDGS-IMac14-2-R225-fm)'. The main window shows a data table with columns for 'LA_Defim_Prod', 'FUA_im_Prod', 'Bsp', 'LA_Notiz', 'Lumping', 'LA_InterneNotizen', 'Syn MB', 'LA_Kommentar', 'LA_Regional', and 'LA_Tiefe'. The table contains several rows of data, each representing a different word sense (Lesart). The interface is heavily annotated with red and blue boxes, arrows, and text, indicating a workflow or process. The top of the screen shows the FileMaker menu bar and a search bar. The bottom of the screen shows a red bar with the number '15'.

Abgrenzen und Beschreiben der Lesarten in FileMaker

Interne Definition

2313 Reihenfolge_im_Pr 1
 LA 10
 Zeit als Ressource: verfügbare oder benötigte Zeit für Tätigkeit, Aufgabe oder Prozess oder Zeit, die zur freien Gestaltung zur Verfügung steht; Zeitraum, Zeitdauer

Definition für Eintrag

LA_Def_im_Prod
 Zeitdauer, die für etwas zur Verfügung steht oder benötigt wird

F UA_im_Produkt -->ÜÄ
 Zeit
 Uhrzeit
 Termin
 Datum

Übersetzungsäquivalente

2322

disambiguierende Zusätze

Äquivalent	Kontextangabe_eintrag	Markierung	in_Eintrag	in index?	abweichende Sortierung	Kontext für Index	reihenfolge
Zeit			<input type="checkbox"/> nein	<input type="checkbox"/> nein		fester Zeitpunkt	1
Uhrzeit			<input type="checkbox"/> nein	<input type="checkbox"/> nein			2
Termin			<input type="checkbox"/> nein	<input type="checkbox"/> nein			3
Datum			<input type="checkbox"/> nein	<input type="checkbox"/> nein			4

typische Mundbilder

MB_P zeit, [MG]

Belege/Beispiele

Bsp gewählt fertig bearb
 > ilex -> ilex Beleg ins WB
 sele Belege aus Korpus
 Bel1 B1 .id=2429002 2
 Bel2 B2 .id=2396882 1
 Bel3 B3 .id=2728261
 Bel4 B4 .id=2116160
 Bel5 B5 .id=2264583

(Teil)synonyme und Antonyme

Syn1	ilex://types.	ä	PHASE2A	[
Syn2	ilex://types.	ä	PHASE1A	[
Syn3	ilex://types.		ZEIT5A	2	
Syn4	ilex://types.	gl	ZEIT7C	2 2 ?	
Syn5					
Ant1					
Ant2					
Ant3					

Sachgruppen

- Zeit
- Lebenslauf
- Geschichte

Kommentar, Muster, Semantisch-Syntaktisches

LA_Kommentar_im_Produkt
 LA_regional
 Form: oft mit mehrfacher Wiederholung, mit Blick auf das Handgelenk und ggf. eingebettet in eine
 LA_typische_Muster -->Muster Belege i >Belege
 ZEIT1 [SCHNELL1A / SCHNELL2 / SCHNELL3A / SCHNELL6]
 LA_Sem_Synt
 DWS_status_LA gesichert (ins Produkt)

2319

Status der LA

interne Bearbeitungsnotizen

zu_besprechen: weiteres Beispiel suchen?	LA_interneNotizen_Bearbeiterin 91 ilex: 4 oft eingebettet in CA, mit blick auf Handgelenk, oft ohne Mundbild, mit intensiv-Mimik, Später: Definition/Formulierung: Drückt aus...
	Probleme: MB, ÜÄ, ; Formkommentar prüfen: Synonym STARTSCHUSS

Arbeitsschritt 9: Korpusbeispiele für die Produktion auswählen und bearbeiten

Abfrage 3

```

null as status, null as eintrag
union
select 2317, LAnr_in_FM.value, 2, 1, workflow.value, workflow2.value, komm.value as kommentar,
case when kontext.value is not null then '[' || kontext.value || ']' else "end",

```

1 von 7 ausgewählt

la_fm	la_ilex	sort	auswahl	workflow1	workflow2	kommentar	kontext	uebers	informant	type	hamnosys	mb	url	url2	status	eintrag
2317	2317	2	1	alles fertig	Schnitt: ✓ Übers...			Im Internet war es eine schöne <... FRA-39	ZEIT1	zeit	ilex://clips.id=3... ilex://tags.id=2...	gesichert (ins P...				354
2317	2317	2	2	zu prüfen (DGS)		[Die Erzählerin fühlte sich im Inter...	Als diese Zeit vorbei wa... STU-47	ZEIT1	zeit	ilex://clips.id=3... ilex://tags.id=2...	gesichert (ins P...					354
2317	2317	2	3	alles fertig	Schnitt: ✓ Übers...	[Thema: Dopingkontrollen beim G...	Früher, zu der Zeit, gab ... KOE-14	ZEIT1	zeit	ilex://clips.id=3... ilex://tags.id=2...	gesichert (ins P...					354
2317	2317	2					Aber wenn man daran denkt, in ... HB-05	ZEIT1	zeit	ilex://clips.id=2... ilex://tags.id=2...	gesichert (ins P...					354
2317	2317	2					Es war eine schöne Zeit! (aus TK... FRA-39	ZEIT1	zeit	ilex://clips.id=2... ilex://tags.id=2...	gesichert (ins P...					354
2317	2317	2					Das war eine total schöne Zeit. (... KOE-11	ZEIT1	zeit	ilex://clips.id=2... ilex://tags.id=2...	gesichert (ins P...					354

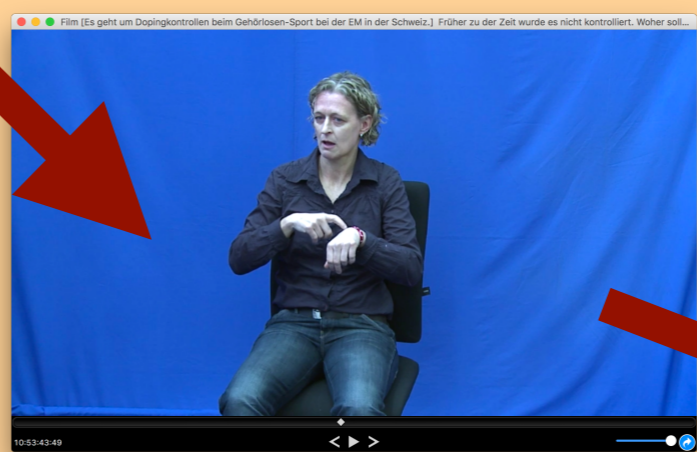
Bsp gewählt fertig bearb

> ilex -> ilex b Beleg ins WB

sele Belege aus Korpus

Bel1	B1	.id=2364366	1
Bel2	B2	.id=2347941	
Bel3	B3	.id=2309304	
Bel4	B4	.id=2363700	2
Bel5	B5	.id=2368551	3

2317



Timecode

Timecode	Deutsche Ü...	Englische Ü...	Segmentieru...	Lexem/Gebä...	HamNoSys_A	Beispiel_A	Komme
10:53:42:00							
10:53:42:11							
10:53:42:20							
10:53:42:22							
10:53:42:29	In der Vergangenheit wurde das nicht kontrolliert.	They didn't do these tests in the past.					Früher zu der Zeit wurde es nicht kontrolliert. Woher sollte das Geld kommen?
10:53:42:30							
10:53:42:34							
10:53:42:44							
10:53:42:44							
10:53:42:46							
10:53:42:49							
10:53:43:23							
10:53:43:32							
10:53:43:35							
10:53:43:41							
10:53:43:49							
10:53:44:10							
10:53:44:14							
10:53:44:18							
10:53:44:19							
10:53:44:19							
10:53:44:28							
10:53:44:31	Wo sollte das Geld dafür herkommen?	Where would the money have come from?					
10:53:44:35							
10:53:44:37							
10:53:44:46							
10:53:44:48							
10:53:45:01							
10:53:45:01							
10:53:45:07							
10:53:45:07							
10:53:45:09							
10:53:45:14							
10:53:45:14							
10:53:45:22							
10:53:45:22							
10:53:45:32							
10:53:45:38	Ab den Wettspielen in Dänemark, 1997	Since the World Games in Denmark, 1997					
10:53:45:38	in Dänemark, ab dem wurden die Dopingkontrollen verpflichtend.	Denmark, the tests against stimulant usage were mandatory, verpflichtend.					
10:53:45:43							
10:53:45:46							

Ausschnitt und Übersetzung angepasst

10:38:11:49 - 10:38:12:08

Tokens Analy Standbilder Feedback-Verw

Rechte Hand

Glosse: ZEIT1

HamNoSys: [d\ə...]

mod:

HamNoSys:

dir:

lok:

Bedeutung:

Pluralität: Nicht zutreffend

Interpretation:

Linke Hand

Abfrage 8

```

select distinct case when md6.value is not null then '[' || md6.value || ']' else "end" || case when beispiel.value is null then uebers.value || ' (angepasst für WB: ilex://clips.id=' || beispiel.id || ']' when beispiel.value is null then uebers.value || ' (aus TK: ilex://clips.id=' ||

```

1 Einträge

?column?	tagsid
[Es geht um Dopingkontrollen beim Gehörlosen-Sport bei der EM in der Schweiz.] Früher zu der Zeit wurde es nicht kontrolliert. Woher sollte das Geld kommen? (angepasst für WB: ilex://clips.id=3151081)	3151081



Arbeitsschritte: 11a. Regielistenerstellung & 11b. Studioaufnahmen

Regieliste aus FMP für copy & paste

Informationen für Regieliste (hier Einzelgebärdenaufnahmen): aus FMP-Einträgen generiert nach Abgleich mit schon vorhandenen Filmen

FMP

Regieliste:

1. Glossenliste aktualisieren
2. Studiofilmliste aktualisieren
3. Gesamtcheck
- 4 relevante DS aufrufen, ggf. sortieren
5. zu Regielistenlayout

Regieliste Einträge

-->Liste Muster

```
113ztabzvar1ztabzMISCHEN4-$SAMztabz?????ztabzøztabzztabzztabz13150
119ztabzzus4ztabzBEZAHLEN8ztabz?????ztabzbezahlztabzztabzztabz51332
119ztabzzus6ztabzFINANZIERUNG2ztabz?????ztabzfinanzztabzztabzztabz7916
119ztabzla_694ztabzBEZAHLEN1ztabz?????ztabzbezahlztabzztabzztabz8513
123ztabzvar1ztabzTIER4-$SAMztabz?????ztabzøztabzztabzztabz15846
125ztabzvar1ztabzGESICHT1-$SAMztabz?????ztabzøztabzztabzztabz15400
140ztabzvar1ztabzBAUM2-$SAMztabz?????ztabzøztabzztabzztabz18032
```

Nachbearbeitung

```
MISCHEN4-$SAM  ?????? ø
BEZAHLEN8  ?????? bezahl
FINANZIERUNG2  ?????? finanz
BEZAHLEN1  ?????? bezahl
TIER4-$SAM  ?????? ø
GESICHT1-$SAM  ?????? ø
BAUM2-$SAM
????? ø
```

```
MISCHEN4-$SAM
?????
ø
BEZAHLEN8
?????
bezahl
FINANZIERUNG2
?????
finanz
BEZAHLEN1
?????
bezahl
TIER4-$SAM
?????
ø
GESICHT1-$SAM
?????
ø
BAUM2-$SAM
?????
ø
```

Im Studio

MISCHEN4-\$SAM

Ⓞ² 110 0+

ø

06.04.20 2

BEZAHLEN8

Ⓞ² 110 0

bezahl

06.04.20 3

FINANZIERUNG2

Ⓞ² 110 0

finanz

06.04.20 4

BEZAHLEN1

Ⓞ² 110 0

bezahl

06.04.20 5

#	Eintrag	LA	Glosse	HNS	MB	Komm	Protokoll
113	var1		MISCHEN4-\$SAM	Ⓞ² 110 0+	ø		
119	zus4		BEZAHLEN8	Ⓞ² 110 0	bezahl		
119	zus6		FINANZIERUNG2	Ⓞ² 110 0	finanz		
119	la_69		BEZAHLEN1	Ⓞ² 110 0	bezahl		
123	var1		TIER4-\$SAM	Ⓞ² 110 0+	ø		
125	var1		GESICHT1-\$SAM	Ⓞ² 110 0	ø		
140	var1		BAUM2-\$SAM	Ⓞ² 110 0	ø		