

# Auswahl von Aufnahmen für das Teilkorpus

Bearbeitet von: Uta Salden, Reiner Konrad  
Stand: 2014-07-07

## 1 Ziele des Teilkorpus

Das Teilkorpus soll die verschiedenen Facetten der DGS zeigen. Die dafür ausgewählten DGS-Texte sollen daher möglichst vielseitig und abwechslungsreich sein. Es wurde angestrebt, dass das Teilkorpus

- die regionale Vielfalt und die verschiedenen Altersgruppen abbildet,
- unterschiedliche Aufgaben der Erhebung enthält,<sup>1</sup>
- eine große Vielfalt an unterschiedlichen Themen abdeckt,
- verschiedene Gebärdenstile und möglichst alle Informanten zeigt.

Im Folgenden wird das Vorgehen im Einzelnen beschrieben.

## 2 Auswahl der im Teilkorpus enthaltenen Aufgaben (Tasks)

Um eine vorläufige Prioritätenliste für die Zusammensetzung des Teilkorpus zu erstellen, wurden die Projektmitarbeiterinnen und -mitarbeiter gebeten einzuschätzen, welche Aufgaben ihrer Meinung nach von besonderer Wichtigkeit und besonderem Interesse für die Gehörlosengemeinschaft sind.

	<b>Relevanz der Aufgaben (Tasks)</b> 1 = nicht aufnehmen, 2 = kann wegfallen, 3 = aufnehmen (25-50 % pro Altersgruppe, 4 = aufnehmen (50-75 % pro Altersgruppe, 5 = aufnehmen (100 % pro Altersgruppe); MA = Mitarbeiter								
<b>Aufgabe (Task)</b>	<b>MA 1</b>	<b>MA 2</b>	<b>MA 3</b>	<b>MA 4</b>	<b>MA 5</b>	<b>MA 6</b>	<b>MA 7</b>	<b>MA 8</b>	<b>Wert (Ø)</b>
Erfahrungen als Gehörloser	4	4	5	5	5	3	5	5	4,5
Witz	5	5	3	5	3	3	5	3	4,0
Freie Konversation	4	4	4	4	4	4	3	3	3,8
Diskussionsthemen	5	4	3	4	4	4	3	3	3,8
Sachgebiete	4	3	3	3	3	4	4	4	3,5
Erlebnisbericht	3	3	3	4	4	3	4	4	3,5
Heimatregion	4	4	3	3	3	3	3	3	3,3
Gehörlosenveranstaltungen	3	3	4	3	3	3	4	3	3,3
Einzelgebärden	5	1	1	2	4	5	3	3	3,0
Junge vs. alte Gebärden	2	1	3	3	3	3	5	3	2,9
<i>Katze und Vogel (7 Teile)</i>	3	2	2	3	3	3	3	3	2,8
<i>Frosch und Junge (6 Teile)</i>	3	2	2	3	1	3	3	3	2,5
Birnengeschichte	3	2	2	3	1	3	3	3	2,5

<sup>1</sup> Siehe Zwischenbericht 2009 und 2010.

Termine	3	2	1	3	3	4	1	2	2,4
Ablaufbeschreibung	2	2	2	3	3	3	2	2	2,4
Schilder	2	2	1	3	3	3	2	2	2,3
Vater und Sohn	2	2	3	1	2	2	2	2	2,0
<i>Reisegeschichte (7 Teile)</i>	2	1	2	3	2	3	1	2	2,0
Lustige Geschichte	2	2	2	2	1	1	3	3	2,0
Feueralarmgeschichte	1	2	3	2	2	1	1	2	1,8
Namensgebärden	2	1	1	2	1	2	2	2	1,6
Stadtspaziergang	1	1	1	2	2	1	1	1	1,3
Realer Film (Signs)	1	2	1	1	2	1	1	1	1,3

Daraufhin wurde den ersten acht Aufgaben – mit Ausnahme der Aufgabe „Witz“ – in der Prioritätenliste entsprechend ihrem Rang ein Anteil an den 50 Stunden des Teilkorpus zugewiesen.

Aufgabe (Task)	Anteil an Gesamtstunden (%)	Gesamtzeit (Std.)
Erfahrungen als Gehörloser	20	10
Freie Konversation	15	7,5
Diskussionsthemen	15	7,5
Sachgebiete	10	5,00
Erlebnisbericht	10	5,00
Heimatregion	7,5	3,75
Gehörlosenveranstaltungen	7,5	3,75
<i>Zwischensumme</i>	<b>85,0</b>	<b>42,5</b>

Die Aufgabe „Witz“ wurde als wichtig erachtet, da sie eine hohe kulturelle Bedeutung hat. Da die Witze im Durchschnitt nur 1,67 Minuten lang sind, kann der Aufgabe „Witz“ im Teilkorpus kein vergleichbar großer Anteil an den Gesamtstunden zugeordnet werden wie den anderen Aufgaben mit hoher Relevanz. Die Festlegung der Auswahlkriterien und die Auswahl dieser Aufgabe wurden auf 2014 verschoben.

Die Aufgaben, denen eine niedrigere Relevanz zugeordnet wurde, wurden aus den folgenden Gründen nur exemplarisch für das Teilkorpus ausgewählt. Lediglich die Aufgabe „Einzelgebärden“ wurde gar nicht ausgewählt.

Aufgabe (Task)	Relevanz	Status	Begründung
Einzelgebärden	3,0	nicht ausgewählt	sind eher von lexikographischem Interesse (Varianz) und dienen nicht den Zielen des Teilkorpus
Junge vs. alte Gebärden	2,9	exemplarisch	experimentelle Erhebungsmethode; kann zu spezifischer Forschung (metalinguistische Sprachproduktion) genutzt werden
Katze und Vogel (7 Teile)	2,8	exemplarisch	ähnlicher Output bei den Informanten, daher nur exemplarisch ausgewählt; kann zur crosslinguistischen Forschung genutzt werden (produktive Gebärden, Constructed Action, nonmanuelle Komponenten)
Frosch und Junge (6 Teile)	2,5	exemplarisch	(s. Katze und Vogel)
Birnengeschichte	2,5	exemplarisch	(s. Katze und Vogel)
Termine	2,4	exemplarisch	ähnlicher Output bei den Informanten; wird zur Einarbeitung der studentischen Mitarbeiter genutzt, daher sind etliche Tasks bereits

			transkribiert
Ablaufbeschreibung	2,4	exemplarisch pro Thema	Bandbreite der Themen soll gezeigt werden; bei Wahl desselben Themas ist der Output der Informanten ähnlich; kann zu Forschungszwecken genutzt werden (Textsorten, -strukturierung)
Schilder	2,3	exemplarisch	ähnlicher Output bei den Informanten; wird zur Einarbeitung der studentischen Mitarbeiter genutzt, daher sind etliche Tasks bereits transkribiert
Vater und Sohn	2,0	exemplarisch	ähnlicher Output bei den Informanten; kann zu Forschungszwecken genutzt werden (Nacherzählung nach unterschiedlichen Stimuli, hier: Bildergeschichte; siehe auch Feuealarmgeschichte und Lustige Geschichte)
Reisegeschichte (7 Teile)	2,0	exemplarisch	ähnlicher Output bei den Informanten; kann zu spezifischer Forschung (Modifikation von sog. Richtungsgebärden) genutzt werden
Lustige Geschichte	2,0	exemplarisch	ähnlicher Output bei den Informanten; kann zu Forschungszwecken genutzt werden (Nacherzählung nach unterschiedlichen Stimuli, hier: Filmszene in DGS; siehe auch Vater und Sohn und Feuealarmgeschichte)
Feuealarmgeschichte	1,8	exemplarisch	ähnlicher Output bei den Informanten; kann zu Forschungszwecken genutzt werden (Nacherzählung nach unterschiedlichen Stimuli, hier: Erzählung in DGS; siehe auch Vater und Sohn und Lustige Geschichte)
Namensgebärden	1,6	exemplarisch	kann zu spezifischer Forschung genutzt werden
Stadtspaziergang	1,3	exemplarisch	Da das Stimulus-Material bereits im Projekt <i>Dicta Sign</i> genutzt wurde, stehen bereits DGS-Texte dieser Art zur Verfügung.
Realer Film (Signs)	1,3	exemplarisch	ähnlicher Output bei den Informanten; kann zu Forschungszwecken genutzt werden (Diskussion über einen Film ohne gesprochene Sprache)

### 3 Auswahl der im Teilkorpus enthaltenen Subtasks

Nachdem die Anteile der einzelnen Aufgaben am Teilkorpus festgelegt wurden, musste entschieden werden, nach welchen Kriterien die einzelnen Subtasks innerhalb einer Aufgabe ausgewählt werden sollten. Generell wurde für das Teilkorpus entschieden, dass jeder Informant mindestens einmal und möglichst aktiv, d.h. als Gebärdende/Gebärdender, ausgewählt wird und die Faktoren Geschlecht und Alter der Informanten möglichst ausgeglichen sind. Für die Gewichtung der Regionen waren zwei Vorgehensweisen denkbar:

1. Eine repräsentative Auswahl: Der Informantenanteil einer bestimmten Region entspricht dem Zeitanteil am Teilkorpus. (Beispiel: x % aller Informanten kommen aus Region A, daher enthält das Teilkorpus für eine bestimmte Aufgabe x % des Materials von Informanten aus Region A). Dies hat zur Folge, dass aus Erhebungsregionen mit nur 16 Informanten (Mindestanzahl) nur wenig Material pro Aufgabe ausgewählt wird. Für Personen aus diesen Regionen könnte das Material weniger interessant sein.
2. Eine regional gefärbte Auswahl: Für eine bestimmte Aufgabe werden aus jeder Region gleich viele Informanten ausgewählt. Eine solche Auswahl hat zur Folge, dass aus Erhebungsregionen mit nur 16 Informanten einige Informanten doppelt ausgewählt

werden, aus Erhebungsregionen mit einer hohen Informantenzahl dagegen einige Informanten nicht ausgewählt werden.

Um die Vor- und Nachteile der beiden Alternativen aufzuwiegen, wurde eine Kombination festgelegt: Zunächst werden wie unter 2. aus jeder Region pro Aufgabe gleich viele Informanten ausgewählt, solange bis in einer Region alle Informanten einmal ausgewählt wurden. Danach wird pro Aufgabe wie bei 1. ausgewählt, jedoch mindestens ein Subtask aus jeder Region.

Es ist davon auszugehen, dass ein Subtask nicht komplett für die Veröffentlichung geeignet ist, da er auch Gesprächsanteile mit wenig inhaltlichem Wert (Nachfragen etc.) enthält. Aus diesem Grund wurde für die Kalkulation die durchschnittliche Subtask-Länge um 15% gekürzt. Zusätzlich wurde anhand des festgelegten prozentualen Anteils an den 50 Stunden des Teilkorpus der reale Zeitanteil in Stunden errechnet. Mit diesen beiden Werten wurde kalkuliert, wie viele Subtasks einer Aufgabe etwa im Teilkorpus enthalten sein werden.

Aufgabe (Task)	Länge Subtasks (Ø in min)	-15 %	Anteil an Gesamtstunden (%)	Gesamtzeit (Std.)	Anzahl Subtasks
Erfahrungen als Gehörloser	11,49	9,76	20	10	61,4
Diskussionsthemen	18,56	15,78	15	7,5	28,5
Freie Konversation	16,19	13,76	15	7,5	32,7
Witz	1,67		2,9	1,44	52,0
Sachgebiete	12,89	10,96	10	5,00	27,4
Heimatregion	19,19	16,31	7,5	3,75	13,8
Gehörlosenveranstaltungen	9,29	7,90	7,5	3,75	28,5
Erlebnisbericht	7,46	6,34	10	5,00	47,3

Wie oben bereits erwähnt, soll im Teilkorpus möglichst abwechslungsreiches Material gezeigt werden. Mit Aufgaben, bei denen durch das Stimulus-Material unterschiedliche Themen vorgegeben waren, wurde daher die Auswahl begonnen. Die Anwendung der Kriterien für den jeweiligen Subtask wird im Folgenden beschrieben.

### 3.1 Sachgebiete

Insgesamt wurden 26 verschiedene Themen als Stimuli vorbereitet. Bei jeder Erhebung wurden vier Themen angeboten, von denen die Informanten zwei auswählen sollten. Durch die Zusammenstellung der angebotenen Themen wurde bereits bei der Erhebung gesteuert, dass der Abstand zwischen beliebten und daher häufig ausgewählten und weniger beliebten Themen nicht zu groß wird. Mit Ausnahme des Sachgebiets DDR (acht mal) wurde jedes Sachgebiet mindestens zwölfmal ausgewählt.<sup>2</sup> Nach der obigen Tabelle sollten 27 Subtasks ausgewählt werden (entsprechend dem prozentualen Anteil an der Gesamtlänge des Teilkorpus sowie der durchschnittlichen Länge der Subtasks dieses Tasks). Es wurden folgende Kriterien festgelegt:

- Jedes Thema wird einmal ausgewählt.
- Aus jeder Region werden zwei Subtasks ausgewählt.
- Alter und Geschlecht sind innerhalb der Aufgabe möglichst ausgeglichen.

Entsprechend wurden 26 Subtasks ausgewählt.

<sup>2</sup> S. Zwischenbericht 2012, Projektinternes Arbeitspapier AP02-2012-01.

### 3.2 Erlebnisberichte

Insgesamt wurden 14 verschiedene Themen erhoben. Den Informanten wurde eins von zwei Sets mit jeweils 7 Themen angeboten, aus dem sie ein Thema auswählen konnten. So wurden die beiden Sets zwar gleich vielen Informanten angeboten, durch die freie Auswahl innerhalb der Sets wurden jedoch einige Themen häufig, andere selten ausgewählt. Laut Tabelle sollten 47 Erlebnisberichte ausgewählt werden. Dadurch ergaben sich folgende Kriterien:

- Jedes Thema sollte möglichst einmal ausgewählt werden, die übermäßige Auswahl eines Themas sollte vermieden werden.
- Aus jeder Region werden 4 Subtasks ausgewählt.
- Alter und Geschlecht sollten innerhalb einer Region möglichst ausgeglichen sein.
- Es werden bevorzugt Monologe ausgewählt.

Entsprechend sollten 52 Erlebnisberichte ausgewählt werden. Die Sichtung (s. Punkt 4) ergab jedoch, dass ein Subtask ungeeignet war. Daher wurden nur 51 Subtasks ausgewählt.

### 3.3 Diskussionsthemen

Insgesamt wurden 8 verschiedene Themen erhoben. Den Informanten wurde eins von zwei Sets mit jeweils 4 Themen angeboten, aus dem sie ein Thema auswählen konnten. So wurden die beiden Sets zwar gleich vielen Informanten angeboten, durch die freie Auswahl wurden jedoch einige Themen häufig, andere selten ausgewählt. Laut Tabelle sollten 28 Diskussionsthemen ausgewählt werden. Dies ergab folgende Kriterien:

- Jedes Thema sollte möglichst einmal ausgewählt werden, die übermäßige Auswahl eines Themas sollte vermieden werden.
- Aus jeder Region werden 2 Subtasks ausgewählt.
- Alter und Geschlecht sollten innerhalb der Aufgabe möglichst ausgeglichen sein (auch das Geschlecht innerhalb einer Altersgruppe).

Entsprechend sollten 26 Subtasks ausgewählt werden. Die Sichtung (s. Punkt 4) ergab jedoch, dass ein Subtask ungeeignet war. Daher wurden nur 25 Subtasks ausgewählt.

### 3.4 Gehörlosenveranstaltungen

Den Informanten wurden mehrere Veranstaltungen als Anregung vorgeschlagen, sie konnten jedoch auch über eine andere Veranstaltung berichten, die nicht vorgeschlagen wurde. Laut Tabelle sollten 29 Subtasks ausgewählt werden. Folgende Kriterien wurden festgelegt:

- Jedes Thema sollte möglichst einmal ausgewählt werden, die übermäßige Auswahl eines Themas sollte vermieden werden.
- Aus jeder Region werden 2 Subtasks ausgewählt.
- Alter und Geschlecht sollten innerhalb der Aufgabe möglichst ausgeglichen sein (auch das Geschlecht innerhalb einer Altersgruppe).

Entsprechend wurden 26 Subtasks ausgewählt.

Nach Abschluss der Auswahl der Gehörlosenveranstaltungen waren in den kleinen Erhebungsorten fast alle Informanten einmal ausgewählt worden. Ab der folgenden Aufgabe wird die Auswahl daher wie auf Seite 3 unter 1. beschrieben prozentual vorgenommen.

### 3.5 Erfahrungen als Gehörloser

Die Informanten konnten frei zum Thema „Erfahrungen als Gehörloser“ erzählen, ohne dass Stimulusmaterial vorgegeben war. Laut der Tabelle sollen 61 Erfahrungsberichte ausgewählt werden, die Anzahl pro Region ergibt sich aus dem Anteil der Informanten aus der Region bezogen auf das Gesamtkorpus. Bei den Erfahrungsberichten werden überwiegend Monologe ausgewählt. Die Anzahl der ausgewählten Subtasks stimmt dann überein mit der Anzahl der

ausgewählten Informanten. Für Erhebungsorte, bei denen 6 oder weniger Informanten noch nicht ausgewählt wurden (goe, hb, hh,.mvp, nue, sh), wurde festgelegt, dass nur 1 oder 2 Informanten für die Aufgabe „Erfahrungen als Gehörloser“ ausgewählt werden sollten, um die restlichen, nicht ausgewählten Informanten besser auf die weiteren Aufgaben verteilen zu können. In diesen Regionen wurden folglich weniger Informanten ausgewählt, als es der prozentuale Anteil der Region am Gesamtkorpus zugelassen hätte. Die Zahl der Informanten aus den kleinen Regionen, die nicht ausgewählt wurden, wurde auf die größeren Regionen umverteilt. Die folgende Tabelle zeigt die Zahlen im Einzelnen.

	noch nicht ausgewählte Informanten	Informantenanteil der Region am Gesamtkorpus (%)	Anzahl <sup>3</sup>	korrigierte Anzahl	Differenz	Zuschlag	endgültige Anzahl für Auswahl
<b>ber</b>	13	8,48	5			1	<b>6</b>
<b>fra</b>	18	9,70	6			2	<b>8</b>
<b>goe</b>	5	6,06	4	2	-2		<b>2</b>
<b>hb</b>	3	4,85	3	1	-2		<b>1</b>
<b>hh</b>	1	4,85	3	1	-2		<b>1</b>
<b>koe</b>	29	13,33	8			2	<b>10</b>
<b>lei</b>	15	9,09	6			1	<b>7</b>
<b>mst</b>	19	9,70	6			2	<b>8</b>
<b>mue</b>	12	7,88	5			1	<b>6</b>
<b>mvp</b>	3	4,85	3	1	-2		<b>1</b>
<b>nue</b>	4	5,45	3	2	-1		<b>2</b>
<b>sh</b>	2	4,85	3	1	-2		<b>1</b>
<b>stu</b>	21	10,91	7			2	<b>9</b>

Zusätzlich wurden folgende Kriterien festgelegt:

- Alter und Geschlecht sollten innerhalb der Aufgabe möglichst ausgeglichen sein (auch das Geschlecht innerhalb einer Altersgruppe). Da schon relativ viele Informanten ausgewählt wurden, ist dieses Kriterium nicht mehr vollständig zu erfüllen.

### 3.6 Freie Konversation

Für diese Aufgabe war kein Stimulusmaterial vorgesehen. Laut Tabelle sollten 32 Subtasks ausgewählt werden, die Anzahl pro Region ergibt sich aus dem Anteil der Informanten aus der Region bezogen auf den Gesamtkorpus. Für Erhebungsorte mit weniger als 4 Informanten, die noch nicht ausgewählt wurden (goe, hb, hh,.mvp, nue, sh), wurde festgelegt, dass nur 1 oder 2 Informanten für die Aufgabe „Freie Konversation“ ausgewählt werden sollten, um die restlichen, nicht ausgewählten Informanten besser auf die Aufgaben verteilen zu können. In diesen Regionen wurden folglich weniger Informanten ausgewählt als der prozentuale Anteil der Region am Gesamtkorpus zugelassen hätte. Die Zahl der Informanten aus den kleinen Regionen, die nicht ausgewählt wurden, wurde auf die größeren Regionen umverteilt. Die folgende Tabelle zeigt die Zahlen im Einzelnen.

<sup>3</sup> Anzahl nach prozentualer Gewichtung.

	noch nicht ausgewählte Informanten	Informantenanteil der Region am Gesamtkorpus (%)	Anzahl <sup>4</sup>	korrigierte Anzahl	Differenz	Zuschlag	endgültige Anzahl für Auswahl
ber	7	8,48	3				3
fra	9	9,70	3			1	4
goe	3	6,06	2	1	-1		1
hb	1	4,85	2	1	-1		1
hh	1	4,85	2	1	-1		1
koe	19	13,33	4			2	6
lei	7	9,09	3				3
mst	11	9,70	3			1	4
mue	5	7,88	2				2
mvp	2	4,85	2	1	-1		1
nue	1	5,45	2	1	-1		1
sh	1	4,85	1				1
stu	9	10,91	3			1	4

Zusätzlich wurden folgende Kriterien festgelegt:

- Alter und Geschlecht sollten möglichst ausgeglichen sein. Dieses Kriterium war auch bei der Aufgabe „Freie Konversation“ nicht vollständig zu erfüllen, da die Auswahl der in Frage kommenden Informanten eingeschränkt war.
- Es sollen möglichst bereits fertig transkribierte Subtasks ausgewählt werden.

Entsprechend wurden 32 Subtasks ausgewählt.

### 3.7 Heimatregion

Für diese Aufgabe war kein Stimulusmaterial vorgesehen. Laut Tabelle sollten 14 Subtasks ausgewählt werden. Folgende Kriterien wurden festgelegt:

- Pro Erhebungsregion wird ein Subtask ausgewählt.
- Alter und Geschlecht sollten möglichst ausgeglichen sein. Dieses Kriterium war aufgrund der fortgeschrittenen Auswahl nur bedingt einzuhalten.

Entsprechend wurden 13 Subtasks ausgewählt.

Die folgende Tabelle zeigt den Stand vom 3.7.2014. Die angestrebten Anteile am geplanten Umfang des Teilkorpus von 50 Stunden wurden mit Ausnahme der Aufgabe „Erfahrungen als Gehörloser“ erreicht. Durch weitere Sichtungsdurchläufe (s. Punkte 4 und 5) wird sich die Gesamtlänge pro Task noch reduzieren, so dass auch die anderen Aufgaben, die exemplarisch ausgewählt werden, noch ausreichend Berücksichtigung finden werden.

Aufgabe (Task)	Anzahl Subtasks ausgewählt	Länge (gesamt)	Anteil an Gesamtstunden (%)	Anteil an Gesamtstunden (geplant; %)
Erfahrungen als Gehörloser	61	12:19:49:44	24,7	20
Freie Konversation	32	08:42:06:14	17,4	15
Diskussionsthemen	25	07:20:02:08	14,7	15
Sachgebiete	26	05:39:08:01	11,3	10
Erlebnisbericht	51	06:12:25:14	12,4	10
Heimatregion	13	04:20:45:40	8,7	7,5
Gehörlosenveranstaltungen	26	03:47:02:49	7,6	7,5

<sup>4</sup> Anzahl nach prozentualer Gewichtung.



<i>Summe</i>		<b>48:21</b>	<b>96,7</b>	<i>85,0</i>
--------------	--	--------------	-------------	-------------

Insgesamt wurden alle Filme einmal ausgewählt, d.h. jeder Informant ist mindestens in einem Subtask zu sehen, wenn auch nicht alle aktiv gebärden. Dies wird bei der noch ausstehenden Auswahl der Aufgabe „Witz“ beachtet, so dass alle Informanten im Teilkorpus mindestens einmal aktiv gebärden.

## 4 Sichtung der ausgewählten Subtasks

Nach der eher statistischen Auswahl der Subtasks soll diese qualitativ in einem ersten Schritt überprüft werden, indem die ausgewählten Subtasks gesichtet werden. Bei der Sichtung werden folgende Aspekte beachtet:

- *Inhalt*: Geeignet für Veröffentlichung? Passt der Inhalt zur Aufgabenstellung? Erzählt der Informant etwas Interessantes zum Thema?
- *Gebärdensstil*: Ist der Gebärdensstil gut zu verstehen? Kann man den Informanten gut transkribieren?
- *Wohlfühl-Faktor*: Ist das Video angenehm anzuschauen? Ist die Beleuchtung in Ordnung? Ist die Haltung der Personen angenehm anzuschauen? Sonstige Besonderheiten?
- *Technische Probleme*: Diese wurden im Zuge der Postproduktion dokumentiert. Sollten technische Probleme vorhanden sein, wird der betroffene Subtask nicht ausgewählt.

Wenn bei der Sichtung Besonderheiten bei einem Informanten festgestellt werden, wird dies in iLex als Metadatum zum Informanten dokumentiert, damit diese Informationen für die weitere Auswahl sowie die Vergabe der Subtasks an die Studierenden im Rahmen der Basistranskription zur Verfügung stehen.

Wenn sich bei der Sichtung herausstellt, dass ein bereits ausgewählter Subtasks nicht für das Teilkorpus geeignet ist, wird dieser aus der Auswahl entfernt und i.d.R. durch einen anderen ersetzt. Dabei wird angestrebt, dass der Ersatz-Subtask denselben Kriterien entspricht (Aufgabenstellung sowie Alter, Geschlecht und Herkunftsregion des Informanten).

Erst nach der Sichtung und dem Austausch von ungeeigneten Subtasks ist die Auswahl für eine Aufgabe abgeschlossen, sodass mit der Auswahl der nächsten Aufgabe begonnen werden kann. So wird vermieden, dass Informanten doppelt ausgewählt werden. Um den Aufwand für die Auswahl zu begrenzen, wurden Subtasks, die durch weitere Bearbeitungs- oder Sichtungsschritte für ungeeignet erachtet wurden, nicht immer ersetzt.

## 5 Übersetzung und weitere qualitative Sichtung

Die ausgewählten Subtasks können nun für das Teilkorpus aufbereitet werden. Im ersten Schritt werden sie übersetzt. Dadurch wird das Material zum ersten Mal inhaltlich erschlossen und ein weiteres Mal gesichtet. Diese Sichtung hat das Ziel, Teile im Subtask zu identifizieren, die nicht geeignet sind für die Veröffentlichung, weil sie z.B. inhaltlich uninteressant sind, nicht zur Aufgabe passen etc. Die zweite Sichtung führt dazu, dass nur Material transkribiert wird, das später auch im Teilkorpus veröffentlicht werden soll. Diese Information wird im Kommentar zum Subtask dokumentiert.