

Segmentierung

Bearbeitet von: Thomas Hanke, Sung-Eun Hong, Susanne König, Reiner Konrad, Gabriele Langer, Silke Matthes, Rie Nishio, Anja Regen

Stand: 2011-07-22 (Reiner Konrad)

Die Frage, wo eine Gebärde beginnt und wo sie endet, ist durchaus nicht trivial. Auch wenn wir grundsätzlich davon ausgehen, dass DGS-kompetente Gebärdende intuitiv eine Gebärde im natürlichen Gebärdenfluss identifizieren können, scheint es uns sinnvoll, die Regeln der Segmentierung festzulegen, um eine möglichst konsistente Erfassung der Tokens zu erreichen. Damit präzisieren wir die bisher am IDGS in anderen Projekten bereits angewandte Praxis und machen sie explizit. Abweichend von anderen Projekten¹ haben wir uns für eine enge Segmentierung entschieden, bei der Übergangsbewegungen nicht als Teil der Gebärde angesehen werden. Dies verdoppelt zwar die Anzahl der zu annotierenden Segmentgrenzen, da Ende einer Gebärde und Beginn der Folgegebärde in der Regel nicht zusammenfallen, man erreicht dadurch aber, dass die Formen der einzelnen Tokens eines Types weniger stark voneinander abweichen, da insbesondere Unterschiede am Ende einer Gebärde wegfallen. Dies führte insbesondere bei einem automatisierten Abgleich der Token-Formen zu weniger Differenzen (visuelles Rauschen), die eine manuelle Bearbeitung erfordern.

Im Folgenden werden zunächst die Regeln der Segmentierung wiedergegeben, die nicht nur im DGS-Korpus-Projekt, sondern auch im Dicta-Sign-Projekt angewendet werden. Anschließend folgen zusätzliche Konventionen für bestimmte Gebärdentypen oder Artikulationsphänomene.

Segmentierungsregeln

Es gibt zwei grundlegende Herangehensweisen, wie man kontinuierliches Gebärden in Einzelgebärden segmentiert:

- Eine Gebärde beginnt da, wo die vorhergehende Gebärde endet – flüssiges Gebärden bedeutet, dass es keine Lücke zwischen Gebärden gibt.
- Übergangsbewegungen zwischen Gebärden zählen nicht als Teil der Gebärden. Folglich gibt es Lücken zwischen zwei Gebärden, in denen die Artikulatoren vom Ende einer Gebärde zum Beginn der nächsten Gebärde bewegt werden. (In einigen Fällen gibt es keine Übergangsbewegungen und folglich auch keine Lücken.)

Beide Herangehensweisen haben ihre Vor- und Nachteile. Aber z.B. im Rahmen des Dicta-Sign-Projekts bietet die zweite Herangehensweise Vorteile für die weitere Verarbeitung der Daten:

- Ein Token-Tag repräsentiert dann den Teil des Signals, der durch HamNoSys beschrieben ist.
- Die Variation zwischen Tokens ist viel kleiner, als wenn die Übergangsbewegung Teil der Gebärde wäre.

Die meisten Forscher, die diese Herangehensweise befolgen, haben zwar irgendwelche Regeln auf Papier, aber meistens verlassen sie sich auf ihre Intuition, wo geschnitten werden soll. Um Unter-

¹ Unter anderen zur Australischen Gebärdensprache (Auslan): *Endangered Languages Documentation Project* (URL: <http://www.auslan.org.au/about/corpus/>; Transkriptionskonventionen: Johnston, Trevor. 2011: Guidelines for annotation of the video data in the Auslan Corpus. [Manuskript in ständiger Überarbeitung, Stand: 19. 6. 2011; aktuelle Version online im Internet; URL: <http://www.auslan.org.au/about/annotations/>), zur Niederländischen Gebärdensprache (NGT): *NGT Corpus Project* (URL: <http://www.ru.nl/corpusngt/>; Transkriptionskonventionen: Crasborn, Onno / Zwitserlood, Inge. 2008: Annotation of the video data in the Corpus NGT. Nijmegen: Radboud University. [Online im Internet; URL: http://www.ru.nl/aspx/download.aspx?File=/contents/pages/515436/corpusngt_annotationconventions.pdf]), zur Britischen Gebärdensprache (BSL): *BSL Corpus Project* (URL: <http://www.bslcorpusproject.org/>; Transkriptionskonventionen: s. Johnston 2011).

schiede zwischen den unterschiedlichen Projektgruppen möglichst gering zu halten, versuchen wir im Folgenden, diese Regeln so genau wie möglich zu formulieren:

Einhändige Gebärden

Der einfache Fall sind Gebärden mit einer HMH²-Struktur im Sinne von L&J³: Die Gebärde beginnt mit einem Stopp und endet mit einem Stopp. (Wir weisen an dieser Stelle darauf hin, dass dieses Vorgehen streng genommen nicht dem Bottom-up-Ansatz entspricht, d.h. es ist nicht daten-gesteuert, denn wir verwenden unser Wissen über den Type, um die Token zu schneiden/taggen.)

Beginn einer Gebärde:

- Die Gebärde beginnt, sobald ihre Handform vollständig ist und sie sich mit der richtigen Orientierung an der Anfangslokation der Gebärde befindet.
- In Fällen, in denen sich zwei Gebärden einen Stopp teilen (d.h. eine Gebärde endet mit einem Stopp und die nächste Gebärde beginnt zufällig mit einem Stopp an genau derselben Lokation mit derselben Handform und Orientierung), schneide den Stopp in der Mitte. (In diesem Fall ist klar, dass es keine Lücke zwischen den beiden Tags gibt.)
- In Fällen, in denen die Gebärden keine spezifischen Anfangslokationen haben, suche eine Unterbrechung der Bewegung (z.B. eine plötzliche Richtungsveränderung) vom Ende der vorhergehenden Gebärde bis zum Ende der zu schneidenden Gebärde. Dieser Punkt ist der Beginn der Gebärde.
- Im Falle einer kontinuierlichen Bewegung vom Anfang einer Gebärde zum Ende der nächsten Gebärde (z.B. DENKEN DU beim lockeren Gebärden), schneide in der Mitte/am Scheitel des Bewegungsbogens. (Dies ist dann gleichzeitig das Ende der vorhergehenden Gebärde, d.h. es gibt keine Lücke zwischen den beiden Gebärden.)

Ende einer Gebärde:

- Wenn die Gebärde mit einem Stopp endet, endet sie am Ende dieses Stopps (schneide vor der ersten Veränderung eines Parameters).
- Wenn die Gebärde mit einer Bewegung endet, dann schneide vor einer Veränderung der Bewegungsrichtung.
- Wenn es keine Bewegungsveränderung gibt, zeigt ein Wechsel der Handform oder der Orientierung das Ende der Gebärde an.
- In Fällen, in denen es keinen Wechsel der Handform oder der Orientierung gibt, aber eine kontinuierliche Bewegung von der vorhergehenden zu der nachfolgenden Gebärde, endet die Gebärde in der Mitte/am Scheitel dieser Bewegung (s.o.).

Zweihändige Gebärden

Im Prinzip kann man die Kriterien für einhändige Gebärden auf jede der beiden Hände anwenden. Das führt in einigen Fällen dazu, dass die Tags für jede Hand unterschiedliche Start- und Endzeitpunkte haben. Wenn man für die Transkription zwei Spuren vorsieht, ist dies einfach zu handhaben – man braucht jedoch mehr Zeit zum Schneiden. Zudem hat dies Auswirkungen auf die Beantwortung der Frage, was aus kognitiver Sicht eine Gebärde darstellt.

Hat man nur eine Spur – und dies trifft auch auf das Schneiden des Videos zu, denn das wird für die Bildverarbeitung benötigt – muss man ein Kriterium finden, das beide Hände berücksichtigt. Interessanterweise ist die einfachste und konsistenteste Anweisung, um beide Hände in einer Spur zu schneiden, sich auf die dominante/aktive Hand zu konzentrieren und die andere zu ignorieren:

Für zweihändige Gebärde wende die o.g. Anweisung auf die dominante/aktive Hand an.

² Hold-Movement-Hold (Stopp-Bewegung-Stopp)

³ Liddell, Scott K. / Johnson, Robert E. 1989: American Sign Language. The phonological base. In: Sign Language Studies 18, 64, 195-277.

Beschreibung des Movement-Hold-Modells auf Deutsch in: Becker, Claudia. 1997: Zur Struktur der deutschen Gebärdensprache. Trier: Wissenschaftlicher Verlag Trier, S. 81-91.

Anwendung des Modells auf Auslan in: Johnston, Trevor / Schembri, Adam. 2006: Australian sign language (Auslan): An introduction to sign language linguistics. Cambridge: Cambridge University Press, S. 111-116.

Zusätzliche Konventionen

- **Indexikalische Gebärden**
Indexikalische Gebärden sowie Vorkommen konventioneller Gebärden, die auf der indizierenden Bilderzeugungstechnik beruhen, werden immer als Movement-Hold-Einheiten angesehen, d.h. die Bewegung, die mit oder ohne Kontakt endet, ist Teil der Gebärde.
- **Fingeralphabet**
Buchstaben, die fließend hintereinander gefingert werden und sich in der Regel auf ein Wort oder einen Eigennamen beziehen, werden als ein Token segmentiert. Unterbrechungen wie Pausen, die mit dem Ende einer Sinneinheit zusammenfallen, oder wenn der Gebärdende zögert und überlegt, sowie eine deutliche Veränderung der Lokation der Hand sind Gründe, mehrere Token-Tags anzulegen. Im Zweifel sollte eher weniger als mehr Token-Tags angelegt werden.
- **Zahlen**
Im Unterschied zu vorhergehenden Projekten werden im DGS-Korpus-Projekt Zahlen nicht mehr als eine Einheit angesehen, sondern in Gruppen aufgeteilt, deren Elemente systematisch miteinander kombiniert werden können. Dadurch kann gezielt auf die verschiedenen Realisierungen der Elemente jeder Gruppe zugegriffen werden. Die Gruppen sind:
 - Zahlen von 0-12
 - Zahlen von 13-19
 - Zehner-Zahlen ab 20 (20, 30, ... 90)
 - Hunderter-Zahlen (100, 200, ... 900)
 - Tausender-Zahlen (1000, 2000, ... 9000)

Wird die Zahl 1925 in DGS gebärdet, dann kann dies in vier Einzel-Token segmentiert werden, entweder $19+100+5+20$ oder $1000+900+5+20$.
- **Gehaltene Gebärden (Unterbrechung des Gebärdenflusses)**
Gebärden, die auffallend länger als normal gehalten werden, weil der Gebärdende z.B. innehält und überlegt – meist verändert sich der Blick in dieser Phase, z.B. schaut der Gebärdende nochmals auf den Monitor, auf seinen Gesprächspartner oder irgendwo hin – oder damit signalisiert, dass er den Turn noch behalten will, dass er noch etwas gebärden will, was ihm momentan gerade nicht einfällt, werden nicht als ein Token segmentiert, sondern aufgeteilt in zwei Segmente. Das erste Segment endet nach den ersten 5 Frames, in denen die Gebärde gehalten wird (bei Filmen mit einer Bildfrequenz von 25 Bildern pro Sekunde (25 fps); bei Filmen mit einer Bildfrequenz von 50 Bildern pro Sekunde (50 fps) sind es 10 Frames). Das zweite Token-Tag endet mit der Veränderung der Gebärde (s. Segmentierungsregel).
Die Dauer von 5 Frames ist willkürlich gesetzt und wird nach einiger Zeit der Praxisanwendung überprüft. Durch das Aufteilen dieser gehaltenen Gebärden in zwei Token-Tags wird erreicht, dass eine Untersuchung der durchschnittlichen Dauer eines Token-Tags durch solche Vorkommen nicht verfälscht wird.
- **Wiederholte, versetzte Bewegungen**
Insbesondere bei sogenannten Richtungsgebärden kann es zu Problemen der Segmentierung kommen, wenn z.B. eine Gebärde wie GEBEN wiederholt und seitlich versetzt ausgeführt wird. Wird die Gebärde jedes Mal vollständig ausgeführt, d.h. geht die Hand immer wieder zum Ausgangspunkt zurück, dann handelt es sich um mehrere Tokens, jedes in der Bedeutung „einer (bestimmten) Person etwas geben“. Ist die Bewegung verschliffen, so dass die Hand nicht jedes Mal zum Ausgangspunkt zurückkehrt, dann handelt es sich um ein Token in der Bedeutung „mehreren (verschiedenen) Personen nacheinander etwas geben“ (Distributiv).
- **Backchannel-Verhalten**
Als Backchannel-Verhalten oder -Signale werden Gebärden, Mimik, Mundbilder und Gestik verstanden, durch die der Gesprächspartner signalisiert, dass er aktiv den Äußerungen des Sprechers/Gebärdenden folgt. Typische Backchannel-Gebärden sind z.B. JA und

STIMMT. Diese Äußerungen des Gesprächspartners werden in der Basistranskripton nicht segmentiert.

Segmentiert werden jedoch Gebärden oder mimisch-gestisches Verhalten, das

- zum Wechsel des Sprechers führt (Turn-Taking),
 - eine Reaktion oder (inhaltliche) Antwort beim Sprecher hervorruft,
 - zu einem Dialog mit einem möglichen Themenwechsel führt.
- **Mundbild**
Die Regeln der Segmentierung berücksichtigen nur die Gebärden, d.h. die manuellen Zeichen. Simultan mit den Gebärden artikulierte Mundbilder (Ablesewörter) haben auf die Segmentierung keinen Einfluss. Geht die Artikulation des Ableseworts über das Ende der Gebärde hinaus oder beginnt ein Mundbild vor der Gebärde, dann ändert dies nichts bei der Bestimmung der Segmentgrenzen.
Mundbilder werden im Rahmen der Basistranskription in erster Linie wegen ihrer Relevanz für die Dokumentation der Bedeutung einer Gebärde oder Gebärdensequenz erfasst. Wir gehen davon aus, dass Ablesewörter in der Regel zeitgleich mit der Gebärde artikuliert werden, wohl wissend, dass im Normalfall Beginn und Ende von Gebärde und Mundbild nicht exakt übereinstimmen.