

Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks *so*

Anja Stukenbrock
(FRIAS, Universität Freiburg/Br.)

Abstract	1
1. Einleitung	2
2. Kategorisierungsfragen: <i>so</i> als deiktisches Sprachzeichen	3
3. Fälle deiktisch-gestischen Gebrauchs von <i>so</i>	5
3.1. <i>so</i> sieht / sehen X aus + Geste	5
3.2. <i>so</i> X (X = Adjektiv) + Geste	9
3.3. <i>so</i> X (X = Verb) + Geste	13
3.4. <i>so</i> + Ø + Geste/Performanz/Inszenierung	15
4. Schlussdiskussion.....	19
5. Literaturverzeichnis.....	21

Abstract

Auch wenn der Ausdruck *so* eine hochfrequente Form im gesprochenen Deutschen darstellt, scheint er sich hartnäckig allen Versuchen einer grammatisch stichhaltigen Klassifikation zu widersetzen. Er veranlasst Linguisten und Linguistinnen entweder dazu, auf einer sehr abstrakten Ebene einen letzten Rest semantischer und/oder funktionaler Gemeinsamkeit herauszudestillieren, um am Konstrukt einer grammatischen Einheit festhalten zu können, oder motiviert sie umgekehrt dazu, ihr linguistisches Objektiv mit einem höheren Auflösungsgrad auf die empirische Wirklichkeit zu richten und ihre Datenbefunde in eine stärker differenzierende Theoriediskussion rückzuführen.

Der Beitrag knüpft an Untersuchungen an, die ausgehend von einer empirisch fundierten Beschreibung unterschiedlicher Vorkommens- und Verwendungsweisen des Ausdrucks *so* für ein kategoriell differenzierteres Verständnis seiner heterogenen Erscheinungsformen plädieren. In meiner Analyse geht es um eine eng umgrenzte Gruppe von *so*-Instanzen, die sich dadurch auszeichnen, dass ihr Verständnis untrennbar mit dem situativ emergierenden Interaktionskontext verbunden ist, in dem die Teilnehmer einen wechselseitigen audio-visuellen Wahrnehmungszugang zueinander haben. Die These lautet, dass in diesen Fällen Projektionsverhältnisse bestehen, die nicht allein auf der Ebene des Verbalen (insbesondere im Bereich der Syntax) zwischen Elementen derselben Modalität vorkommen, sondern die sich durch ihre multimodale, ressourcenübergreifende und diese miteinander vernetzende Qualität auszeichnen. Die Analyse verfolgt Ziel, die komplexen Projektionsverhältnisse zwischen verbalen und visuellen Ausdrucksressourcen beim deiktisch-gestischen Gebrauch des Ausdrucks *so* darzulegen. Als Ergebnis werden vier verschiedene Konstruktionstypen ermittelt, die sich durch die Art und Weise unterscheiden, wie das Deiktikon *so* in den verbalen Kontext eingebettet ist. Allen vier Typen ist jedoch gemeinsam, dass das Deiktikon *so* als Kontextualisierungshinweis auf eine vom Sprecher ausgeführte und vom Adressaten visuell wahrzunehmende Handlung fungiert.

1. Einleitung

Auch wenn der Ausdruck *so* eine hochfrequente Form im gesprochenen Deutschen darstellt, scheint er sich hartnäckig allen Versuchen einer grammatisch stichhaltigen Klassifikation zu widersetzen. Konrad Ehlich bezeichnet ihn im Untertitel seines Aufsatzes zu „*so* – Überlegungen zum Verhältnis sprachlicher Formen und sprachlichen Handelns“ als „widerspenstige[s] Beispiel“ und führt ihn als „Kandidaten für eine geradezu klassische Unschärfe der Kategorisierung“ an (Ehlich 2007: 147). Dieser Kandidat veranlasst Linguisten und Linguistinnen entweder dazu, auf einer sehr abstrakten Ebene einen letzten Rest semantischer und/oder funktionaler Gemeinsamkeit herauszudestillieren, um am Konstrukt einer grammatischen Einheit festhalten zu können (Sandig 1987; Weinrich 1993), oder motiviert sie umgekehrt dazu, ihr linguistisches Objektiv mit einem höheren Auflösungsgrad auf die empirische Wirklichkeit zu richten und ihre Datenbefunde in eine stärker differenzierende Theoriediskussion rückzuführen (Auer 2006; Hennig 2006).

Mathilde Hennig hat in ihrer korpusbasierten Revision der Wortartenklassifikation von *so* nachgewiesen, dass die in unterschiedlichen Nachschlagewerken anzutreffenden Wortartenhypothesen keinesfalls alle Verwendungsweisen von *so* abdecken (Hennig 2006: 427). Hennigs Analyse der Wortartenzugehörigkeit von *so* illustriert zwar das Problem einer Diskrepanz zwischen Theorie und Empirie, doch auch diese Untersuchung erweist sich ungeachtet des Bemühens um einen „enge[n] Konnex von Theorie und Empirie“ (Hennig 2006: 428) als methodisch ergänzungsbedürftig. Gerade bei der Analyse gesprochensprachlicher Daten ist die Beschränkung auf eine quantitative Auswertung schriftlich fixierter Korpora unzureichend. Ohne eine Berücksichtigung vokalsprachlicher Merkmale wie der Prosodie einerseits und körperlich-visueller Aspekte der face-to-face-Interaktion andererseits wird eine Untersuchung von *so* in der mündlichen Kommunikation kategoriell unscharf bleiben bzw. simplifizieren. Daher teile ich Auers grundsätzliche Kritik an bisherigen Kategorisierungsversuchen, die aufgrund der Abstraktion von Vorkommnissen im konkreten Sprachgebrauch verkennen, dass man es mit einer „Vielzahl von wesentlich spezifischeren Konstruktionsschemata“ zu tun hat, „die eigene syntaktische, semantische und pragmatische Eigenschaften haben“ (Auer 2006: 109). Auer untersucht aus konstruktionsgrammatischer Perspektive diejenigen *so*-Konstruktionen, die eine satzwertige Struktur im Folgesyntagma projizieren, und gelangt allein in dieser Gruppe zu sechs verschiedenen Typen, für die postuliert wird, dass sie einzeln gespeicherte und prozessierte Konstruktionen darstellen.

Mein Beitrag knüpft an Untersuchungen an, die ausgehend von einer empirisch fundierten Beschreibung unterschiedlicher Vorkommens- und Verwendungsweisen des Ausdrucks *so* für ein kategoriell differenzierteres Verständnis der heterogenen Erscheinungsformen von *so* plädieren. Anders als Auer geht es in meiner Analyse nicht um syntaktisch projizierende *so*-Konstruktionen, sondern um eine eng umgrenzte Gruppe von *so*-Instanzen, die sich dadurch auszeichnen, dass ihr Verständnis untrennbar mit dem situativ emergierenden Interaktionskontext verbunden ist, in dem die Teilnehmer einen wechselseitigen audio-visuellen Wahrnehmungszugang zueinander haben. In diesen Fällen – so meine These – bestehen Projektionsverhältnisse, die nicht allein auf der Ebene des Verbalen (insbesondere im Bereich der Syntax) zwischen Elementen derselben Modalität vorkommen, sondern die sich durch ihre multimodale, ressourcenübergreifende und diese miteinander vernetzende Qualität auszeichnen. Projektionen fasse ich im Anschluss an Auer (2000, 2005, 2006), Mondada (2006) und Streeck (1995) als grundlegendes Organisationsmerkmal sowohl der Grammatik als auch der Interaktion auf. Im Fall des deiktisch gebrauchten *so* kommen Mikroprojektionen ins Spiel, bei denen es nicht ausreicht, von strukturellen Homologien zwischen Projektionen in Interaktion und Grammatik auszugehen. Statt dessen gilt es, sie als holistisch in einem Paket zusammengeschnürte Einheiten aufzufassen. Gegenstand dieses Beitrags ist daher die Art und Wei-

se, in der körperlich-visuelle Ausdrucksmittel in die *online*-Konstruktion von Turns, Turnkonstruktionseinheiten und deren Einzelkomponenten integriert werden.

2. Kategorisierungsfragen: *so* als deiktisches Sprachzeichen

In einschlägigen Grammatiken, Wörterbüchern und Einzeluntersuchungen wird verschiedentlich auf eine deiktische Verwendungsweise des Sprachzeichens „*so*“ hingewiesen. Im Metzler-Lexikon Sprache, das systematisch die drei klassischen Dimensionen der Personal-, Lokal- und Temporaldeixis aufführt, heißt es, „[s]peziellere Deixis beziehen sich z.B. auf Eigenschaften (dt. *so*)“. Zifonun (1997: 2335) betrachtet das Sprachzeichen *so* in Vergleichssätzen (Beispiel: *Sie singt so gut, wie sie immer gesungen hat*) als deiktisch verweisenden Bestandteil einer Gradphrase. Den weiten Deixisbegriff, der dieser Kategorisierung zugrunde liegt, teile ich nicht und unterscheide mit Lyons (1977) u.a. zwischen Deixis und Phorik. Das von Zifonun gewählte Beispiel fällt daher m.E. unter die Kategorie des kataphorischen Korrelats. Thurmair (2001) untersucht die Funktion von *so* in Vergleichskontexten und setzt diese mit anderen Verwendungsweisen in Beziehung, um zu einer übergeordneten Bedeutung zu gelangen. Ihre Behauptung, dass sich für „alle Verwendungen von *so* [...] feststellen [lässt], daß sie deiktisch sind“ (Thurmair 2001: 27), ist nicht aufrecht zu erhalten (vgl. zur Kritik daran Auer 2006). Dasselbe gilt für ihre Analyse des von einer Geste begleiteten *so* in ad-adjektivischem Gebrauch, dem sie dieselbe Funktion zuschreibt wie dem *so* im expliziten Vergleichskontext mit *wie*-Gruppe (Thurmair 2001: 30). Betrachtet man allein die prosodischen Verhältnisse, werden die Unterschiede (unakzentuiert in der *wie*-Gruppe versus akzentuiert in deiktisch-gestischer Verwendung) deutlich. Ein weiteres Argument gegen eine solche Subsumption liefert die *online*-Syntax (Auer 2000): Während das deiktisch-gestisch verwendete *so* keine auf das Adjektiv folgenden sprachlichen Elemente projiziert, liegt aus der Perspektive einer inkrementellen Syntax die entscheidende Beziehung zwischen dem *so* und der *wie*-Gruppe in der Vergleichskonstruktion darin, dass das *so* strukturaufbauend und die *wie*-Gruppe strukturarbeitend operiert.¹ Zuletzt schließlich ist aus deixistheoretischer Perspektive festzuhalten, dass ein kategorialer Unterschied zwischen textuellen Verweisbeziehungen, die im Fall von *so* + Adjektiv + *wie*-Gruppe m.E. treffender als syntaktische Projektionsbeziehungen aufzufassen sind, und Verweisen auf den außersprachlichen Kontext besteht. In den zuletzt genannten Fällen fungiert der Ausdruck *so* als „Zeigzeichen“ (Bühler 1934/65) auf das situativ emergierende Interaktionsgeschehen und inkorporiert eine visuell wahrzunehmende gestische Handlung in die grammatische Struktur der Rede (Streeck 2002: 582):

So can serve as a ‘flag’ that alerts the interlocutor that there is extralinguistic meaning to be found and taken into account in making sense of what is being said right now. Gaze direction combines with the flag by serving as a pointer to the location where additional meaning is found. In their combination, *so* and the gaze shift serve to direct the interlocutor’s visual attention to the hands and incorporate the work of the hands into the grammatical structure of the talk.

Die von Streeck dargestellte Funktion von *so* im multimodalen Zusammenspiel mit Blick und Gestik wird auch von Fricke (2007) gesehen. Sie schließt zwar die Modaldeixis aus den Dimensionen der Deixis aus, hält jedoch an der Klassifikation des Ausdrucks *so* als Deiktikon fest, mit der Besonderheit, dass sie es unter die Lokaldeixis einordnet (Fricke 2007: 78). Ihre

1 Vgl. zu den Begriffen Auer (2006: 293): „Projektionen sind die Basis der inkrementellen Syntax. Der Wechsel zwischen strukturaufbauenden (Projektionsbögen herstellenden) und strukturarbeitenden (Projektionen einlösenden, Projektionsbögen abschließenden) Phasen konstituiert einen kognitiven Rhythmus, der die Prozessierung erleichtert, indem er den strukturarbeitenden Phasen Entlastungen für andere mentale und interaktive Aktivitäten schafft [...]“

Argumentation gründet sich erstens auf eine Ähnlichkeit zwischen dem Gebrauch von *so/solch* und dem Gebrauch des Demonstrativpronomens *dieser*. Ihr zweites Argument lautet, dass bestimmte Typen des *so*-Gebrauchs eine Zeigegeste erfordern, Zeigegesten sonst aber nur in der Lokaldeixis vorkämen. Eine Ausdifferenzierung erfolgt jedoch nicht, so dass beispielsweise ad-nominale, ad-adjektivische, ad-adverbiale und ad-verbale Verwendungsweisen ebenso wenig unterschieden werden wie die Formvarianten *so* und *solch*, die unzutreffenderweise als Einheit konzeptualisiert werden. Dabei geht der grundsätzliche Funktionsunterschied zwischen dem Gebrauch von *so* als Deiktikon gegenüber *so* als Indexikalitätsmarker (Auer 1981) verloren.

Sehr allgemein fasst Weinrich die Merkmale des Sprachzeichens *so* unter dem Begriff der „Bedeutungsrahmung“ zusammen. Innerhalb der Kategorie der „Status-Adverbien“ (1993: 582), die dazu dienen, einen Sachverhalt auf verschiedene Aspekte wie Form, Qualität, Quantität oder Informationswert einzustellen (ebd.), konstituiert es die Subkategorie „Rahmen-Adverb“ (ebd.: 583). Demnach „rahmt [es] die Bedeutung eines Verbs, Adjektivs oder Adverbs ein“ (ebd.). Die unter diese Kategorie subsumierten Fälle erweisen sich bei eingehender Betrachtung einer größeren Datenmenge als wesentlich heterogener, so dass Differenzierungen erforderlich sind. Ich beschränke mich dabei auf die von Weinrich angeführten Fälle mit Situationsbezug. So nennt er zum einen den Fall, in dem „eine – gegebenenfalls längere – Handlungssequenz in einen Rahmen gestellt und dadurch abgeschlossen werden [kann], so dass gleichzeitig ein neuer Ansatz möglich ist“ (ebd.). Zum anderen führt er den Fall auf, in dem sich das Adverb „auf ein Element der Situation beziehen [kann], das dann eventuell durch eine ikonische Zeigegeste in seinen Konturen umrissen wird“ (ebd.: 584). Das von ihm gewählte Beispiel (*das neue Haus sieht ungefähr so aus*), dem er „eine begleitende Zeigegeste etwa eines spitzgiebligen Hauses“ zuschreibt, macht deutlich, dass eine Zeigegeste nicht nur eventuell erfolgen kann, sondern zwangsläufig erfolgen muss, damit die Äußerung verständlich wird.

Die Gemeinsamkeit zwischen den von Weinrich aufgeführten Fällen besteht darin, dass der Ausdruck *so* in beiden Fällen auf etwas außerhalb der Sprache Wahrnehmbares verweist und dem Adressaten anzeigt, im situativen Kontext nach Verständnishilfen zu suchen. Doch während im letzten Fall eine genaue visuelle Wahrnehmung der körperlich-gestischen Aktivitäten des Sprechers zum sprachlichen Verständnis erforderlich ist, verlangt der erstere Fall nicht notwendigerweise ein präzises Monitoring der Einzelgeste, dafür aber ein globales Situationsverständnis, Ablaufwissen und damit einhergehend Erwartungen über Teilnehmerkonstellationen und mit dem Aktivitätswechsel ggf. einhergehende Veränderungen im *participation framework* (Goffman 1981). Anders als Ehlich (1987/2007: 161f.), der das in dieser Funktion auftretende *so* als „paraexpeditiv“ auffasst und ihm treffend „Funktion eines Scharniers innerhalb der Handlungsabwicklung“ zuschreibt, möchte ich es als *change of activity*-Marker bezeichnen. Es ist nicht deiktisch im engeren Sinn und spielt daher im weiteren Verlauf meiner Darstellung keine Rolle. Das Hauptaugenmerk gilt jenen Verwendungsformen, in denen der Ausdruck *so* eine vom Adressaten wahrzunehmende körperliche Aktivität kontextualisiert, die von einer einfachen Arm-/Handgeste bis zu einer gesamtkörperlichen Demonstration reichen kann.

Es handelt sich also um Fälle, in denen *so* als „Zeigewort“ im „Zeigfeld“ der Sprache fungiert und für seine „Bedeutungserfüllung [...] an sinnliche Zeighilfen gebunden“ ist (Bühler 1934/1965: 80). Der Bühlerschen Definition folgend können sie daher als genuin deiktisch bezeichnet werden. Eine Zeigegeste ist zum Verständnis dieser Gruppe von *so*-Ausdrücken obligatorisch. Ehlich (1987/2007: 152) konstatiert in diesem Zusammenhang: „Dieser Ausdruck wird als Deixis für jeweils spezifische Aspekte von ‚Objekten‘ im Verweisraum eingesetzt.“ Anhand des konstruierten Beispiels *Ich hätte gern fünf Nägel, ungefähr so lang* hebt er die Notwendigkeit der „gestischen Verfahrensweise“ als „Besonderheit“ hervor, „die der deiktische Ausdruck *so* mit dem deiktischen Ausdruck *dieser* und seinen Parallelausdrücken

teilt“ (Ehlich 1987/2007: 153). Andere Autoren sprechen in diesem Zusammenhang von „Zeiggestennotwendigkeit“ (Harweg 1990) oder „Zeigegestenpflicht“ (Sennholz 1985). In der Terminologie Fillmores, der innerhalb der Kategorie deiktischer Ausdrücke erstens nicht-deiktischen von deiktischem Gebrauch und innerhalb des deiktischen Gebrauchs zweitens symbolischen von gestischem Gebrauch unterscheidet, gehören sie zu den gestisch gebrauchten deiktischen Ausdrücken: „By the *gestural* use of a deictic expression I mean that use by which it can be properly interpreted only by somebody who is monitoring some physical aspect of the communication situation“ (Fillmore 1997: 62).

3. Fälle deiktisch-gestischen Gebrauchs von *so*

Im Folgenden werden ausschließlich Fälle diskutiert, in denen der deiktische Gebrauch von *so* mit redebegleitenden Gesten und/oder körperlichen Handlungen korreliert, die zum Verständnis des Modaldeiktikons unabdingbar sind. Von den 30 insgesamt untersuchten Fällen sind alle Instanzen des *so*-Gebrauchs prosodisch durch einen Nukleusakzent gekennzeichnet. Zwischenfälle, in denen das Modaladverb lediglich einen Nebenakzent trägt oder durch Längung prosodisch hervorgehoben ist, werden nicht behandelt. Sie sind nicht nur aufgrund ihrer differnten prosodischen Struktur problematisch, sondern stellen korrelierend mit der weniger starken prosodischen Markierung auch aus gestischer Sicht Zweifelsfälle dar, die eine gesonderte Untersuchung verlangen. So werden zwar in den meisten Fällen redebegleitende Gesten produziert, diese besitzen jedoch nicht denselben Status – obligatorisch – wie die durch das prosodisch akzentuierte, deiktisch-gestische *so* kontextualisierten Gesten. Meine Analyse beschränkt sich auf die prosodisch und gestisch eindeutigen Fälle, deren multimodales Format in den empirischen Analysen zunächst exemplarisch herausgearbeitet und im nächsten Schritt theoretisch reflektiert werden soll.

Datengrundlage der folgenden Analyse bilden drei verschiedene Subkorpora: erstens ein aus Mitschnitten der ersten „Big Brother“-Staffel bestehendes Video-Korpus, zweitens ein Kochkorpus, das Videoaufnahmen zwei verschiedener Sendeformate („Kerners Köche“ und „Die Küchenschlacht“) umfasst, und drittens ein Video-Korpus aus sechs sog. „Schmerzkonferenzen“, auf denen Patienten mit chronischen Schmerzen einem interdisziplinären Ärztegremium ihre Beschwerden vorstellen. Da aus Datenschutzgründen die Abbildung von Patienten aus den Schmerzkonferenzen nicht möglich ist, beschränke ich mich in den Fallanalysen auf exemplarische Fälle aus den anderen Teilkorpora.

3.1. *so sieht / sehen X aus + Geste*

Als semantisch-pragmatische Paradefälle können diejenigen Instanzen betrachtet werden, in denen mittels des Wahrnehmungsverbs *sehen* explizit ausformuliert wird, dass ein durch X Bezeichnetes in seiner Ganzheit gezeigt und visuell wahrgenommen werden soll. Die Funktion des Ausdrucks *so* besteht in dieser Konstruktion darin, den Wahrnehmungsakt des Adressaten temporal an den Äußerungsakt des Sprechers zu binden und ihn mit der Anweisung an den Adressaten zu versehen, das Objekt in seinem Sosein, seiner Phänomenologie wahrzunehmen. Ich möchte sie als *visuelle Evidenz*-Konstruktion bezeichnen und als multimodales Gesamtpaket (*package*; vgl. Goodwin 2003, Heath 1986) konzeptualisieren, das nur durch den Einbezug körperlich-visueller Ausdrucksressourcen adäquat beschrieben werden kann. Zwei Beispielanalysen, von denen die erste den *default-case* und die zweite eine humorvolle Variante davon erfasst, sollen dies verdeutlichen.

Im ersten Beispiel aus dem Datenset „Kerners Köche“ stellt das zu betrachtende X eine spezielle Zutat – die sog. Tonkabohne – dar, die nicht nur selten und daher verhältnismäßig unbekannt, sondern aufgrund ihrer suchtfördernden Inhaltsstoffe in Deutschland sogar verboten ist. Die Köchin (K5) hat jedoch einige Tonkabohnen als Anschauungsobjekte von einem Freund aus Portugal mitgebracht.

Beispiel 1: Tonkabohnen

1 K5: ein (.) lieber FREUND,
 2 grüße an PORTugal,
 3 =an JENS,
 4 hat mir die MITgegeben,
 -> 5 SO schau die tonkabohnen aus;
 6 RIECH=a=ma:l,

Unterschiedliche sinnliche Wahrnehmungen konstituieren den Erfahrungseindruck einer Speisezutat, der den Fernsehzuschauern allerdings nur medial vermittelt werden kann. Zum Erfassen der Eigenschaften der Tonkabohne gehört im Kontext einer Kochsendung nicht nur die visuelle, sondern auch die olfaktorische und insbesondere die gustatorische Wahrnehmung durch die unmittelbar am Geschehen Beteiligten. Da die Geschmacksprobe aufgrund des in Deutschland herrschenden Verbots in diesem Fall ausgeschlossen ist, bleiben lediglich der visuelle Eindruck und die Geruchsprobe, die von der Köchin (K5) durch die Wahrnehmungsverben *ausschauen* und *riechen* bezeichnet und mit der Aufforderung verbunden werden, die entsprechenden Perzeptionshandlungen durchzuführen. Während die zweite dieser Aufforderungen explizit im expeditiven Feld bzw. Lenkfeld der Sprache (Zifonun et al. 1997, Bd. 1: 321) operiert, da sie im Imperativ formuliert und persönlich an den Moderator adressiert wird (Z. 6: *RIECH=a=ma:l*), liegt der Aufforderungscharakter der vorausgehenden Turnkonstruktionseinheit (Z. 5: *SO schau die tonkabohnen aus*;) im multimodalen Gesamtformat der Äußerung und richtet sich nicht allein an den unmittelbaren Interaktionspartner, sondern auch an die Studio- und Fernsehzuschauer. Die folgende Analyse beschränkt sich auf die verbalen und nonverbalen Aktivitäten zur Herstellung der visuellen Wahrnehmungshandlung, da allein sie mit dem in Frage stehenden Modaldeiktikon *so* korreliert.

Zunächst erwähnt die Köchin (K5) den Spender der Tonkabohnen (Z. 1: *ein (.) lieber FREUND*) und winkt, während sie als syntaktischen Einschub einen Gruß an ihn formuliert (Z. 2/3: *grüße an PORTugal, =an JENS,*), mit der linken Hand in die Kamera (Abbildung 1). Parallel dazu greift sie mit der rechten Hand nach den Tonkabohnen auf dem Kochtisch. Dabei lehnt sie sich mit dem Oberkörper nach vorne und wendet dem Moderator Kerner ihren Rücken zu. Anschließend formuliert sie den in Z. 1 mit der Nominalphrase begonnenen und durch den Einschub unterbrochenen Satz weiter (Z. 4: *hat mir die MITgegeben,*). Dabei nimmt sie ihre ursprüngliche Position am Kochtisch wieder ein und richtet sich mit dem Oberkörper erneut auf Kerner aus. Die Wiederherstellung der face-to-face-Orientierung in Form eines *L-arrangement* (Kendon 1990: 213) zwischen ihr und Kerner bildet die Voraussetzung für die Fortsetzung der durch den Fernsehgruß unterbrochenen fokussierten Interaktion.

Simultan zur Artikulation der Anapher *die* in der Verbalphrase (Z. 4: *hat mir die MITgegeben,*) richtet die Köchin ihren Blick ostentativ auf ihre Hände, die nun jeweils eine Tonkabohne halten. Mit dem artikulatorischen Onset des Modaldeiktikons *SO* (Z. 5) hebt sie parallel ihre rechte und linke Hand mit den vertikal nach oben gehaltenen Tonkabohnen auf Brusthöhe an (Abbildung 3) und führt auf diese Weise eine Präsentativgeste aus, die das Zeigeobjekt Tonkabohne in den Aufmerksamkeitsfokus ihres Interaktionspartners und der sekundären Adressaten (Studiogäste und Fernsehzuschauer) rückt. Wie an der Kopf- und Blickorientierung der Interaktanten zu erkennen ist, ist der visuelle Aufmerksamkeitsfokus sowohl der Zeigenden als auch des Adressaten auf den verbal und gestisch hergestellten gemeinsamen Wahrnehmungsraum und die darin befindlichen Demonstrationsobjekte gerichtet.

Abbildung 1	Abbildung 2	Abbildung 3
		
grüße an PORTugal, =an JENS,	hat mir die MITgegeben,	SO schau die tonka-bohnen aus;

Das deiktische Adverb *so* trägt in dieser Intonationsphrase den Nukleusakzent. Es handelt sich um einen klaren Fall dessen, was in der Typologie Fillmores (1997: 62) als gestischer Gebrauch (*gestural use*) deiktischer Ausdrücke gilt. Durch den Gebrauch des Ausdrucks *so* erteilt der Sprecher dem Adressaten die Anweisung, das simultan stattfindende körperliche Geschehen genau zu beobachten und aus der visuellen Wahrnehmung relevante Einsichten über das betreffende Objekt zu gewinnen. Dabei liegt die Besonderheit des Auftretens von *so* in der Konstruktion *so sieht/sehen X aus + Geste* darin, dass die Notwendigkeit der Herstellung eines objektbezogenen perzeptorischen Fokus durch das visuelle Wahrnehmungsverb explizit ausformuliert wird.

Im nächsten Beispiel handelt es sich um eine humorvolle Variante der Präsentation eines Anschauungsobjekts im multimodalen Format *so sieht/sehen X aus*, in der ironisch mit dem Aspekt der Betrachtungswürdigkeit eines unbekanntem/seltenen Anschauungsobjekts als typischem Vertreter seiner Kategorie gespielt wird. Denn bei dem betrachteten Gegenstand handelt es sich um den nur allzu bekannten Alltagsgegenstand Zigarette. Die Sequenz stammt aus der ersten „Big Brother“-Staffel. Darin bekommt Zlatko zum Trost für die soeben erzählte Geschichte über seine Schulprobleme eine Zigarette geschenkt.

Beispiel 2: Zigaretten

- 1 Adr: zladdi NICHT traurig sein-
- 2 [(guck=ma] für DICH,
- 3 Ker: [zladdi]
- 4 Zlt: ((atmet))
- 5 Ker: ((lacht))
- 6 Zlt: was ist DAS?
- 7 Ker: [((lacht))]
- 8 Adr: [((lacht)) ja] geNAU;
- > 9 Jrg: <<f> SO: sehen zigarEtten aus>;
- 10 haste schOn verGESsen;

Das Verbal-Transkript setzt damit ein, dass Andrea (Adr) Trostworte an Zlatko (Zlt) richtet und ihm mit der Fokussierungsaufforderung *guck=ma* (Z. 2) eine Zigarette reicht. Zlatko nimmt die Zigarette entgegen, atmet hörbar ein und aus (Z. 4) und hält seinen Blick auch nach der Zigarettenübergabe weiterhin auf die Zigarette gerichtet, während Kerstin (Ker) lacht (Z. 5). Die Zigarette hält er dabei auf Brusthöhe vor seinem Oberkörper (Abbildung 4):

Abbildung 4	Abbildung 5	Abbildung 6
 <p data-bbox="204 524 469 560">Ker: ((lacht))</p>	 <p data-bbox="590 524 916 560">Zlk: was ist DAS?</p>	 <p data-bbox="1002 524 1362 595">Jrg: SO: sehen zigarEtten aus;</p>

Zu diesem Zeitpunkt (Abbildung 4) dient Zlatkos Blick nicht mehr der interpersonellen Koordinierung (Deppermann/Schmitt 2007) zwischen ihm und Andrea beim Übergeben und Annehmen der Zigarette, das bereits stattgefunden hat. Indem Zlatko weiterhin starr auf die Zigarette blickt, erfährt sein Blick in dieser die Transaktion überdauernden Gerichtetheit eine Funktionsneuzuweisung. Durch seinen Blick hält Zlatko den visuellen Aufmerksamkeitsfokus auf die Zigarette in seiner Hand aufrecht. Dadurch, dass dies sichtbar für die anderen Beteiligten geschieht, fungiert der Blick den anderen gegenüber als interaktives Display und konstituiert einen gemeinsamen Wahrnehmungsraum für weitere Handlungen. Er operiert als nonverbales Zeigzeichen (Bühler 1934/1965) auf einen Wahrnehmungsraum, der projizierend für weitere Handlungen interaktiv relevant gesetzt wird. Die durch den Blick aufgebaute Projektion wird verbal durch die Frage *was ist DAS?* (Z. 6) eingelöst. Der Kontextualisierungshinweis (Auer 1986; Auer/di Luzio 1992) zum Auffinden des Zeigeziels und zur Herstellung der Referenz für das deiktische Demonstrativum *das* erfolgt durch die eingefrorene Blickorientierung als Zeiginstrument auf Suchraum² und Zeigeziel (Abbildung 5).

Auf Zlatkos gespieltes Erstaunen über den vermeintlich fremden, kategoriell nicht einzuordnenden Gegenstand in seiner Hand reagiert Jürgen, indem er als zweites Paarteil zu Zlatkos Frage eine Antwort im multimodalen Format *SO sieht/sehen X aus* + *Geste* produziert. Seine Geste wird als Präsentativgeste in der *palm up open hand*-Form (Kendon/Versante 2003; Kendon 2004) realisiert und von der Blickausrichtung des Sprechers (Jrg) auf das Zeigobjekt in Zlatkos Hand begleitet (Abbildung 6). Die Tatsache, dass Jürgen nicht das syntaktische Format der Frage (*was ist das*) aufgreift und mit der Präsentativkonstruktion *das ist X* antwortet, sondern statt dessen auf die *visuelle Evidenz*-Konstruktion (*so sehen X aus*) rekurriert, ist Ausdruck dessen, dass er Zlatkos Frage nicht als ernst gemeinte Benennungsfrage, sondern als Kontextualisierungshinweis auf ein ironisch-scherzhaftes Spiel mit der (Un)Bekanntheit des Gegenstands verstanden hat. Das Spiel gründet in der gemeinsamen Erfahrung der Bewohner, dass Zigaretten im Big Brother-Haus ein knappes Gut sind, die man so selten zu Gesicht bekommt, dass der Wiedererkennungseffekt erlischt und die kategoriale Wahrnehmung eines entsprechenden Exemplars nicht mehr möglich scheint. Linguistisch betrachtet funktioniert die spielerische Inszenierung der Wahrnehmungswürdigkeit einer Zigarette als exemplarische Vertreterin ihrer Kategorie aufgrund des gemeinsamen sprachlichen Wissens der Beteiligten um den in Beispiel 1 analysierten *default case*. So stellt das Sprachspiel zwischen Jürgen und Zlatko eine ironische Pervertierung des Standardfalls dar, der als Folie für die Ironisierung dient und dadurch als emisch relevante, mental verankerte und jederzeit abrufbare multimodale Einheit im kommunikativen Haushalt der Sprecher ausgewiesen wird.

Es trübe die Sache nicht, diese Konstruktion gleichzusetzen mit einem ähnlichen Fall, in dem das Demonstrativum *das* in der Konstruktion *das ist /das sind X* + *Präsentativgeste* ge-

2 Vgl. zur Unterscheidung von Verweis- und Suchraum Stukenbrock (2009).

braucht wird, und beide unter die Kategorie einer hybriden Lokaldeixis (Fricke 2007) zu subsumieren. Denn anders als Demonstrativa und Lokaldeiktika fügt es der Aufforderung, der körperlichen Aktivität des Sprechenden visuell zu folgen und im unmittelbaren Handlungskontext die verbal nicht vermittelten Informationen zu finden, eine qualitative Komponente hinzu. Der entscheidende Unterschied besteht darin, dass im Fall von *so sieht / sehen X aus* die Prototypikalität des gezeigten X fokussiert wird. Dies ist mit der interaktiven Projektion zweier ganz spezieller Anweisungen an den Adressaten verbunden: Diese lauten, dass der präsentierte Gegenstand erstens perzeptorisch als beispielhaftes Anschauungsobjekt wahrgenommen werden soll, von dem zweitens eine kognitive Repräsentation als exemplarische Realisierungsinstanz gebildet werden soll. Im Gegensatz dazu kann es sich im Fall der multimodalen Konstruktion *das ist/sind X + Geste* bei dem präsentierten Objekt auch um ein untypisches, nicht sofort identifizierbares oder deformiertes Exemplar seiner Gattung handeln, was im Fall von *so sieht/sehen X aus + Geste* gerade ausgeschlossen ist. Dieses Ausschlusskriterium ist ein weiterer Kontextualisierungshinweis für den Adressaten, wie er die Zeigehandlung zu verstehen hat: nämlich als eine, die ihm etwas vorführt, das er als prototypisch in seinem Wissensbestand abspeichern kann.

Zusammengefasst fungiert die multimodal zu konzeptualisierende *visuelle Evidenz*-Konstruktion als Anweisung an den Adressaten, seine visuelle Aufmerksamkeit auf das Zeigziel, das präsentierte Objekt zu richten und es in seiner Phänomenologie, im Hinblick auf kategoriebildende, saliente Merkmale wahrzunehmen und als prototypischen Vertreter seiner Kategorie mental abzuspeichern. Anders gestaltet sich der semantisch-pragmatische Zusammenhang in Fällen, in denen nicht einzelne Objekte in ihrer phänomenologischen Ganzheit wahrgenommen, sondern entweder bestimmte Eigenschaften von Objekten (Abschnitt 3.2.) oder prozessuale Abläufe, Verfahren- oder Handlungsweisen (Abschnitte 3.3. und 3.4.) fokussiert werden.

3.2. *so X (X = Adjektiv) + Geste*

Eine weitere gestische Verwendungsweise des Modaldeiktikons *so* tritt in adverbialer Voranstellung zu einem (prädikativ gebrauchten) Adjektiv auf. In diesen Fällen wird durch das Adjektiv an einem Gegenstand X eine äußerlich sichtbare Eigenschaft (Größe, Länge, Höhe, Breite etc.) fokussiert, deren spezifischer Grad bzw. deren Dimensionalität erst durch die Geste angezeigt wird. Semantisch betrachtet handelt es sich um Ausdrücke aus der Klasse der relativen Adjektive (Eisenberg 2006: 240ff.), die ihre aktuelle Bedeutung in Abhängigkeit von ihrem jeweiligen Bezugsgegenstand und damit aus dem Kontext erhalten.

Bei der Verwendung eines relativen Adjektivs zusammen mit dem deiktisch-gestisch gebrauchten Adverb *so* lässt sich darüber hinaus feststellen, dass eine Arbeitsteilung zwischen verbalen und visuellen Ausdrucksressourcen stattfindet: Diese Arbeitsteilung besteht darin, dass die semantisch als relativ zu geltende Eigenschaft verbal bezeichnet wird, während die Dimensionalität bzw. der spezifische Grad, in dem die Eigenschaft auf das Bezugsobjekt zutrifft und es dadurch als durchschnittlich oder abweichend charakterisiert, visuell durch die Geste ausgedrückt wird. Das Modaldeiktikon übt dabei eine „flag function“ (Streeck 2002: 582) aus, indem es als verbaler Kontextualisierungshinweis auf die simultan stattfindende gestische Aktivität des Sprechers verweist. Anders als bei den oben diskutierten Beispielen 1 und 2, in denen ein Objekt in seiner Ganzheit vorgeführt und die Konstruktion *so sieht/sehen X aus* in der Regel von einer Präsentativgeste begleitet wird, gehören die Gesten im vorliegenden Fall zur Klasse der ikonischen Gesten (Kendon 2004: 100; McNeill 1992: 12; McNeill 2005: 39).

Beispiel 3 illustriert diesen Fall anhand einer Kochszene aus der Sendung „Die Küchenschlacht“, in der Laienköche eine Woche lang unter der Aufsicht eines Fernsehkochs gegeneinander antreten und in jeder Sendung von einem weiteren prominenten Sternekoch evaluiert werden. Als Wettbewerbspreis darf der Gruppensieger am Ende zusammen mit bekannten

Fernsehköchen ein Gericht zubereiten. In der folgenden Sequenz vergleicht der moderierende Fernsehkoch (HL) den Zuschnitt des Thunfischsteaks, das einer der Kandidaten (K4) in der laufenden Sendung zubereitet, mit den Dimensionen des Koteletts, das dieser tags zuvor gekocht hat.

Beispiel 3: so dick

```

1      HL:  und ich muss Sagen,
2          (0.3) GEStern,
3          (0.5) dein KOTElett rolf;
4          (0.4) war schon=n KNALLer;
5      K4:  ja;
-> 6      HL:  dat war ja SO dick;
7      K4:  .hhh ja;
8      P:   ((lacht))
9      HL:  a:ber dat THUNfischsteak;
10     P:   ((lacht))
11     K4:  ((lacht))
12          das kann man noch TOPpen;

```

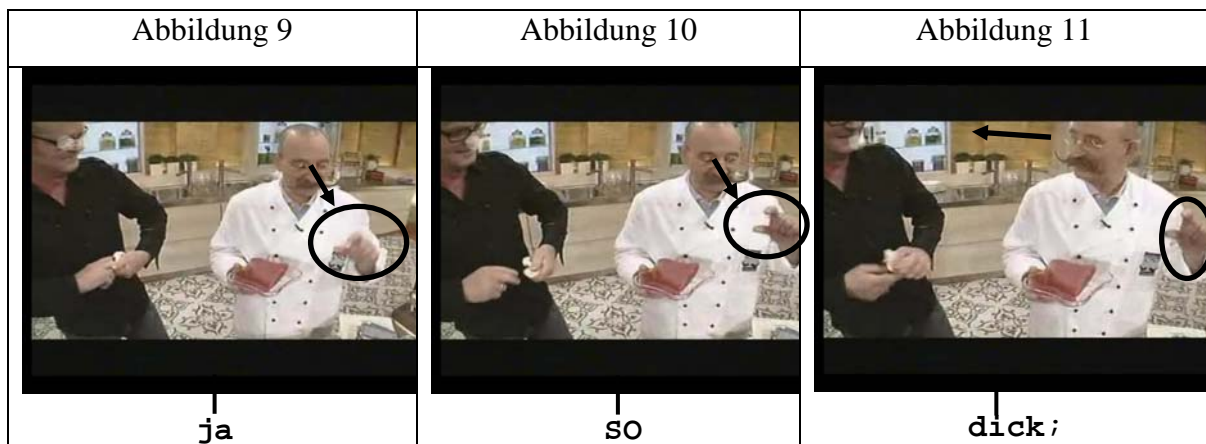
Das Hauptaugenmerk meiner multimodalen Analyse liegt auf der *online*-Emergenz von verbalem und visuellem Geschehen in dem im Verbaltranskript durch die Zeilen 4-6 bezeichneten Teilausschnitt. Im Detail gilt es, durch eine Mikroanalyse die genauen temporalen Relationen zwischen den *online* produzierten verbalen Ausdrücken und Blick- und Gesteneinsatz zu ermitteln und dadurch ressourcenübergreifende Projektionsverhältnisse, d.h. Projektionsverhältnisse zwischen Verbalem, Blick und Gestik zu rekonstruieren.

Der moderierende Fernsehkoch (HL) befindet sich am Kochtisch von Kandidat 4 (K4) und hat das Thunfischsteak, das dieser zubereitet, in die rechte Hand genommen, thematisiert das aktuelle Kochgeschehen jedoch zunächst nicht, sondern ruft mit einem evaluativen Kommentar das Gericht des Vortags in Erinnerung: *GEStern, (0.5) dein KOTElett rolf; (0.4.) war schon=n KNALLer;* (Z. 2-4). Dabei blickt er seinen Adressaten an (Abbildung 7). Dieser reagiert auf den Kommentar mit einem Rückmeldesignal (Z. 5: *ja;*). In welcher Hinsicht das Kotelett bemerkenswert war, erhellt insbesondere für die Zuschauer, denen das gemeinsame Wissen über die vorangegangene Kochsendung nicht zur Verfügung steht, erst aus der nächsten Turnkonstruktionseinheit. Sie spezifiziert die Zuschreibung *KNALLer* (Z. 4) im Hinblick auf den Durchmesser des Kotelettstücks: *dat war ja SO dick;* (Z. 6), die ohne eine begleitende Geste zur ikonischen Repräsentation des Durchmessers vage bleiben würde.

Wichtig im Hinblick auf die temporalen Relationen der Herstellungsaktivitäten des Sprechers (HL) ist die Feststellung, dass dieser bereits beim Onset der Anapher *dat* den Blickkontakt mit seinem Adressaten auflöst und stattdessen in Richtung seiner Hände zu blicken beginnt. Dadurch veranlasst er den Adressaten, seiner Blickorientierung zu folgen und ebenfalls in Richtung des durch den Blick hergestellten Wahrnehmungsraums zu schauen (Abbildung 8).



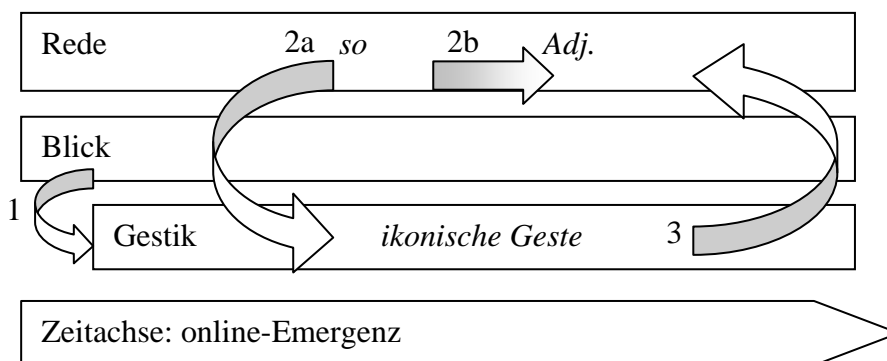
Als nächstes löst sich die linke Hand des Sprechers vom Thunfischsteak und bewegt sich aufwärts. Zugleich beginnt die Hand, die nachfolgende ikonische Geste zu formen, bei der Daumen und Zeigefinger von der restlichen, zur Faust geschlossenen Hand abgespreizt werden. Diese Form wird bereits erkennbar im Verlauf der Artikulation der Partikel *ja* (Abbildung 9). Der Blick beider Interaktanten ist dabei auf die Hand des Sprechers gerichtet. Simultan zur verbalen Produktion des akzentuierten Modaldeiktikons *SO* erreicht die Geste ihren Gipfelpunkt, durch den die vertikale Aufwärtsbewegung des Arms zum Abschluss und die ikonische Fingerformation zur Repräsentation des Kotelettdurchmessers in die exakte Position gebracht wird (Abbildung 10). Noch vor Beendigung der Turnkonstruktionseinheit durch das Adjektiv *dick* löst sich der Blick des Zeigenden wieder vom Zeigeraum und richtet sich erneut auf den Adressaten (Abbildung 11), während die ikonische Geste gehalten und durch einen kurzen rhythmischen Beat visuell akzentuiert wird. Der Beat, dessen temporale Qualität im Standbild nicht abbildbar ist, erfolgt simultan der verbalen Produktion des Adjektivs *dick*.



Wie der weitere Interaktionsverlauf verdeutlicht, dient die Vergegenwärtigung eines abwesenden Gegenstands der vergleichenden Profilierung eines präsenten Gegenstands: des Thunfischsteaks, dessen Zuschnitt das Kotelettstück noch übertrifft und daher vom moderierenden Fernsehkoch als Anlass zu gemeinsamer Erheiterung und zur Publikumsunterhaltung genutzt wird. Bei der deiktisch-gestischen Repräsentation des Kotelettdurchmessers handelt es sich um eine Deixis am Phantasma (Bühler 1934/65: 133ff.). Etwas nicht Präzentes, das Kotelettstück vom Vortag, wird in den Vorstellungsraum der Beteiligten geholt und hinsichtlich einer bestimmten Eigenschaft (überdurchschnittlich dick zu sein) im aktuellen Wahrnehmungsraum durch die ikonische Geste visualisiert.

Aus den temporalen Relationen, die durch die multimodale Mikroanalyse herausgearbeitet wurden, ergeben sich folgende Projektionsverhältnisse zwischen verbalen und visuellen Ausdrucksressourcen: Die simultan zur Turnfortsetzung mit der Anapher *dat* erfolgende Reorientierung des Blicks auf die Hände projiziert den Onset der Gestik (Blick → Gestik). Das akzentuierte Modaldeiktikon *so* projiziert syntaktisch ein Adjektiv X (Rede → Rede), zugleich projiziert es aber auch eine nonverbale Aktivität, genauer gesagt eine ikonische Geste (Rede → Gestik). Die ikonische Geste wiederum projiziert, da sie simultan zum Modaldeiktikon produziert wird, dem Adjektiv hingegen vorausläuft, aufgrund ihrer formellen Gestalt ein semantisch spezifisches Adjektiv (Gestik → Rede).

Streeck (1995: 100) konstatiert bei seiner Analyse der Temporalität und Projektivität ikonischer Gesten folgende Relationen: „Iconic gestures commonly precede their ‚speech-affiliates‘; they unfold – and often decay – before the word arrives. This makes them interesting phenomena for a study of projection in interaction.“ Es stellt sich die Frage, welches im vorliegenden Fall als das sog. *speech-affiliate* der ikonischen Geste anzusehen ist. Für das Adjektiv *dick* spricht die semantische Ko-Expressivität, für das Modaldeiktikon *so* die Ko-Temporalität von Sprachausdruck und Gestenproduktion sowie die funktionale Obligatorik der Geste beim deiktisch-gestischen *so*. Im letzteren Fall trifft die von Streeck behauptete Präzedenz der Geste nicht zu, denn der Gipfel- und Haltepunkt der Geste wird simultan zur Artikulation des Modaldeiktikons erreicht. Im ersteren Fall kann das Argument von der Präzedenz der Geste aufrecht erhalten werden, allerdings ist zu berücksichtigen, dass der Sprecher in meinem Beispiel simultan zur Artikulation des Adjektivs *dick* noch einen auf die ikonische Geste aufgesattelten Beat produziert. Die zusammenfassende schematische Modellierung verdeutlicht, dass sich die Projektionsverhältnisse insgesamt komplexer gestalten, da das Adjektiv doppelt projiziert wird: verbal und gestisch, wobei die verbale Projektion auf der syntaktischen Ebene, die Geste aufgrund ihrer Ikonizität auf der semantisch-repräsentationalen Ebene operiert:



Projektionsbogen 1 veranschaulicht die Projektion des Onsets der Geste durch den vorauslaufenden Blick. Der vom Deiktikon *so* ausgehende Projektionsbogen 2a illustriert die Kontextualisierung der ikonischen Geste; der ebenfalls vom Modaldeiktikon ausgehende Projektionsbogen 2b veranschaulicht die syntaktische Projektion des Adjektivs. Projektionsbogen 3 schließlich verdeutlicht, dass auch von der ikonischen Geste eine auf Formähnlichkeit (Visualisierung des Fleischdurchmessers) beruhende Projektion des Adjektivs stattfindet. Während also das Deiktikon zwei Projektionsbögen aufbaut (2a: gestisch; 2b: syntaktisch), die durch die Geste und durch das Adjektiv eingelöst werden, arbeitet das Adjektiv zwei Projektionsbögen ab (2b: syntaktisch, 3: gestisch), womit sich die Gestalt des multimodalen Gesamtpakets schließt.

3.3. *so X (X = Verb) + Geste*

Mit der Konstruktion *so X (X = Adjektiv) + Geste* ist die Konstruktion *so X + Geste* verwandt, bei der der X-Slot durch ein *Verb* besetzt ist, das eine körperliche Handlung bezeichnet. Dabei wird das Modaldeiktikon dem Verb in der Regel vorangestellt. In seltenen Fällen kann es jedoch auch nach dem Verb stehen. Die Positionsvarianten haben Konsequenzen für das multimodale Realisierungsformat der Gesamthandlung. Als Konstante bleibt jedoch, dass sowohl bei der ad-verbialen Voranstellung als auch bei der Nachstellung der Gebrauch des Ausdrucks *so* exakt denjenigen Moment in der *online*-Emergenz von Sprache und Gestik markiert, in dem die relevante Handlung ausgeführt wird. Meine Analyse beschränkt sich auf die ad-verbiale Voranstellung des Modaldeiktikons, da sie in meinen Daten die weitaus häufigere Variante und damit den Standardfall darstellt.

Im folgenden Beispiel aus der Sendung „Kerners Köche“ erläutert einer der eingeladenen Fernsehköche (K2), wie man idealerweise eine Ananas schält, schneidet und anrichtet. Während der gesamten Erklärung führt der Koch begleitende Bewegungen mit der rechten Hand, die ein Messer hält, und der freien linken Hand aus. Die Analyse beschränkt sich auf die Demonstration der durch das Verb *abschälen* bezeichneten Handlung.

Beispiel 4: *so abschälen*

- 1 K2: und wenn man die **jetzt** mit dem MESSer schneidet,
 -> 2 muss man also SO Abschälen,
 3 muss man SEHR viel wegschneiden,=
 4 = weil diese braunen AUGen,
 5 die (.) INNen noch sind,
 6 müssen RAUS;
 7 Ker ja: ,




Abbildung 12



Zu Beginn seiner Erläuterung (Z. 1: *und wenn man die jetzt mit dem MESSer schneidet*), nimmt der Koch ein Messer in die rechte Hand (Abbildung 12). Der Blick der beiden Interaktanten ist auf den Wahrnehmungsraum gerichtet. Das Temporaldeiktikum *jetzt* markiert den *wenn*-Satz als temporal zu interpretierenden Adverbialsatz und projiziert die Ausführung der durch die Präpositionalphrase bezeichneten Handlung im Sprechzeitraum. Mit Sprechzeitraum ist aufgrund der syntaktischen Projektion des nachfolgenden Hauptsatzes nicht allein die punktuelle *encoding time* (Fillmore 1997: 67) des Adverbialsatzes selbst, sondern die *online*-Emergenz der gesamten, zu diesem Zeitpunkt noch unabgeschlossenen Turnkonstruktionseinheit gemeint.

Während er den Hauptsatz produziert, demonstriert der Koch eine der Teilhandlungen dessen, was er zusammenfassend unter Ananasschneiden versteht. Die temporalen Relationen zwischen der *online*-Emergenz der Rede und der ausführenden Geste gestalten sich wie folgt: Simultan zur Artikulation des Reformulierungsindikators (Gülich/Kotschi 1987: 220; vgl. auch Dittmar in diesem Band) *also*, der auf eine Spezifizierung der im *wenn*-Satz begonnenen Erklärterung verweist, setzt er das Messer am oberen Ende der Ananas an (Abbildung 13) und führt es in einer Abwärtsbewegung außen an der Ananas entlang (Abbildung 14). Diese Abwärtsbewegung vollführt jedoch selbst keinen Schnitt, sondern stellt lediglich eine ikonische Simulation des Schneidens dar. Sie korreliert zeitlich exakt mit der verbalen Produktion

des Modaldeiktikons *so*. Bezeichnenderweise wird der handlungsspezifisierende Verbalinfinitiv *abschälen* erst in der Beendigungsphase der vorgeführten gestischen Handlung artikuliert. D.h. hier bestehen die gleichen temporalen Relationen zwischen vorauslaufender Geste und *speech-affiliate* (Streeck 1995: 100) wie beim Gebrauch von *so X (X = Adjektiv) + Geste*.

Abbildung 13	Abbildung 14	Abbildung 15
		
muss man also	so	Abschälen,

Im vorliegenden Fall handelt es sich nicht nur um eine Erläuterung, sondern zugleich um eine Instruktion, der, wie das Modalverb *muss* signalisiert, im Kontext fachgerechten Kochens eine gewisse deontische Qualität zukommt. Die Funktion des deiktisch-gestischen *so* liegt darin, den entscheidenden Moment in der emergierenden körperlichen Demonstration, dem besondere visuelle Aufmerksamkeit gebührt, zeitlich exakt zu indizieren. Diese spezifische Funktion, zugleich Indikator von etwas visuell Wahrzunehmendem und Indikator des entscheidenden Fokussierungsmoments zu sein, erklärt das gehäufte Auftreten des deiktisch-gestischen Gebrauchs von *so* in Instruktionkontexten, in denen Sprecher neben verbalen Erklärungen körperliche Demonstrationen einsetzen, um ihren Adressaten spezifische Formen von Wissen (prozessuales Wissen, Handlungsabläufe, motorische Fertigkeiten) zu vermitteln.




Noch deutlicher werden die Projektionsverhältnisse zwischen deiktisch-gestisch gebrauchtem *so* in der Funktion als Zeigzeichen auf eine genau wahrzunehmende körperliche Handlung, die erst nachträglich durch ein Handlungsverb deskriptiv bezeichnet wird, im nächsten Beispiel. Es ist interaktionsstrukturell ganz analog zum vorangegangenen Fall aufgebaut. Der Unterschied liegt lediglich in der zeitlichen Zerdehnung, die dadurch zustande kommt, dass es sich bei der zu fokussierenden Handlung um den Essensvorgang handelt, dessen Ausführung die verbale Produktion unterbricht bzw. verzögert. In dieser Sequenz vermittelt der Koch seinen Adressaten, dass Spaghetti legitimerweise „weggeschlabbert“ werden dürfen und führt dies *lege artis* vor.

Beispiel 5: Nudeln wegschlabbern

- 1 K1: und dürfen wir eigentlich nochmal Sagen-
- 2 das wird IMMER gefragt;
- 3 man DARF sie richtig wegziehen; ne?
- 4 man DARF die nudeln wegSCHLABbern und ZIEhen,
- 5 und es DARF auch so=n ganz klein bißchen spritzen;
- 6 ja?
- 7 Ker: [ja]
- 8 K1: [also] im MUND;
- 9 man darF sie RUH:ig so=-
- 10 =WIRKlich;=
- 11 meine damen und herren auch DRAUßen an der-
- 12 an den FERNsehern,
- > 13 SO::-
- 14 (0.4)

- 15 Ker: an den:=o' empFANGSgeräten;
 16 K1: SO:: (.) WEGziehen; (.)
 17 JA;

Auch hier ergeben sich die oben bereits herausgearbeiteten Projektionsverhältnisse zwischen verbalen und visuellen Ausdrucksressourcen. Während der Koch das Fernsehpublikum anspricht, richtet er seinen Blick bereits auf den Wahrnehmungsraum – die mit Spaghetti gefüllte Gabel (Abbildung 16) – und beginnt simultan zur Artikulation des akzentuierten Modaldeiktikons *so*, dessen langgezogene, steigend-fallend-steigende Intonation die Temporalität des Spaghettihochziehens ikonisch abbildet, die Gabel in den Mund zu schieben (Abbildung 17). In der anschließenden 0.4-sekündigen Pause wird das Spaghettischlabbern, begleitet von Schlürfgeräuschen, ausgeführt (Abbildung 18).

Abbildung 16	Abbildung 17	Abbildung 18
		
an den FERN sehern,	SO ::	(0 . 4)

Erst nach Abschluss des Essvorgangs formuliert der Koch noch einmal verbal, wie der Spaghettiverzehr ausgeführt werden soll (nämlich als *WEGziehen*, Z. 16), und greift dabei erneut auf das Modaldeiktikon zurück. Während der erste *so*-Gebrauch katadeiktisch die körperlich-visuelle Handlung projiziert und den Aufmerksamkeitsfokus auf die Aktivitäten im blickorganisatorisch bereits hergestellten Wahrnehmungsraum richtet, schließt der zweite *so*-Gebrauch anadeiktisch den Mikrointeraktionsrahmen, innerhalb dessen die relevante Demonstrationshandlung vollzogen wurde. Auch hier haben wir es mit einer Art Instruktion zu tun (wie man Spaghetti isst), die allerdings weniger deontischen, als vielmehr permissiven Charakter hat. Wie der nächste Konstruktionstyp zeigt, können bei instruierenden Interaktionstypen dieser Art erstens die handlungsbezeichnenden Verben ganz ausgespart und die produzierten multimodalen Formate auch kollaborativ realisiert werden.

3.4. *so* + \emptyset + *Geste/Performanz/Inszenierung*

Bei diesem letzten Konstruktionstyp ist das Modaldeiktikon *so* verbal nicht eingebettet. Die von ihm kontextualisierte körperliche Aktivität stellt in der Regel mehr als eine einzelne Geste dar und umfasst eine gesamtkörperliche Performanz oder Inszenierung des Sprechers. Der Verweisraum für das Modaldeiktikon *so* ist daher der performierende Körper. Er fungiert als *display*-Raum und Medium der dargebotenen Performanz. Ähnlich wie beim Zeigen am eigenen Körper (vgl. Stukenbrock 2008), bei dem der Körper des Zeigenden identisch ist mit dem Körper, auf den gezeigt wird, so dass er zugleich *Subjekt* und *Objekt* der Zeigehandlung ist, stellt auch der Körper des Performierenden zugleich *Performanzsubjekt* und *Performanzobjekt* dar und muss daher als *semiotischer Doppelraum* (ebd.: 8) konstituiert werden. Wie im Folgenden dargelegt werden soll, spielt die Projektionskraft des Modaldeiktikons *so* dabei eine Schlüsselrolle, indem es als verbales Vermittlungsglied zwischen den Herstellungs- und den eigentlichen Performanzhandlungen der jeweiligen Akteure fungiert.

Die nachfolgende Sequenz stammt aus dem „Big Brother“-Korpus. Drei der Bewohner haben sich im Garten versammelt, um Golf zu spielen. Verena hat sich gegenüber ihren Mitbewohnern als Golf-Spezialistin ausgewiesen und beginnt, nachdem der Wechsel zur neuen Aktivität hergestellt ist (Z. 1), John Instruktionen zur richtigen Schlagtechnik zu geben. Insgesamt zwölf Mal wird der Ausdruck *so* in dem Sequenzausschnitt verwendet. Sieben davon stellen Instanzen des deiktisch-gestischen Gebrauchs dar: Z. 7, 9, 12, 14, 16, 26 und 34. Bei den anderen Fällen handelt es sich um den hier nicht untersuchten Gebrauch als *change of activity*-Marker (s.o. Abschnitt 2): Z. 1, als Indexikalitätsmarker (Auer 1981): Z. 23 und 28, sowie als Vorlaufkonstruktion mit *so* (Auer 2006: 298ff.): Z. 22. Da aus Platzgründen nicht alle sieben Instanzen des deiktisch-gestischen *so*-Gebrauchs behandelt werden können, beschränke ich mich in meiner Analyse auf die Fälle in Z. 7, 12 und 16.

Beispiel 6: Golfturnier

```

1   Jhn: [SO: kinder;      ]
2   Ver: [(hab mich) schon] für=s golfturnier VORbereitet;
3   .h nicht wie IHR;
4   (1.0)
5   es=stand in den SPIELregeln;
6   Alx: he=he
-> 7   Jhn: SO: oder wat;
8   Alx: [((hustet))]
-> 9   Ver: [(1.0)      ] SO:?
10  (0.5)
11  ja:;(-)
-> 12  SO:,
13  den HIERhin, .hh
-> 14  SO; ne,
15  (0.8)
-> 16  un' (-) und [dann SO; ]
17  Alx:          [sternchen] ((macht Lockgeräusche))
18  (0.5)
19  Ver: [WEISSte?]
20  Alx: [<<p>komm]=[ma HER;>]
21  Jhn:          [hm=hm,  ]
22  Ver: <<all>also=es=is> SO,
23  des sin' dEr muss (.) hIer muss so (.) FLACH sein;
24  ne?
25  Jhn: hm,
-> 26  Ver: SO.
27  und dann musst du quasi SO;
28  ich will hier nich so:
29  wenn des jemand SIEHT (ja)
30  dann [sagt der nachher]
31  Alx:          [((lacht))      ]
32  Ver: die alte die kt' erzählt (hier n)=SCHEISS .h
33  wenn er RAUSkommt hat er(h)
-> 34  Jhn: SO::: unjefähr;

```

Abbildung 19 veranschaulicht das *participation framework* (Goffman 1981: 137) vor Beginn der Sequenz. Verena, John und Alex befinden sich in einer halbkreisförmigen face-to-face-Konstellation, in der alle denselben *participation status* eines ratifizierten Teilnehmers (Goffman 1981: 131) haben. John übernimmt den Golfschläger von Verena, beendet durch

den *change of activity*-Marker (Z. 1: *SO: kinder;*) das verbale Vorspiel und stellt dadurch den Rahmen für die körperliche Praxis des Golfspielens her. Als nächstes ändert er seine Position im Raum, indem er mit drei Schritten eine Vierteldrehung ausführt, infolge derer er Alex den Rücken zuwendet und sich in face-to-face-Position zu Verena bringt. Dadurch ändert sich das *participation framework* auf grundlegende Weise: Alex verlässt den Interaktionsraum und tritt damit aus der fokussierten Interaktion heraus (Abbildung 20). Das *main involvement* besteht im Folgenden zwischen Verena und John, die eine interaktive Dyade bilden, während Alex sich im weiteren Verlauf der Hauskatze Sternchen zuwendet (Z. 17, 18, 20) und lediglich ein einziges Mal durch das *byplay* (Goffman 1981: 134) eines kommentierenden Lachens (Z. 31) beteiligt.

John leitet nicht nur den Aktivitätswechsel ein und bewirkt durch sein körperliches Verhalten eine Neustrukturierung der Teilnehmerkonstellation, sondern er initiiert durch die Frage *SO: oder wat-* (Z. 7) auch die nachfolgende Instruktionssequenz. Die temporalen Relationen gestalten sich dabei wie folgt: Noch bevor John seine Frage formuliert, hat er mit leicht gegrätschten Beinen, der Positionierung des Schlägers, dem vornüber gebeugten Oberkörper und dem auf den Ball gerichteten Blick (vgl. Abbildung 20) die Haltung eingenommen, die – wie seine anschließende Frage verdeutlicht – von seiner Interaktionspartnerin wahrgenommen und ggf. korrigiert werden soll.


Abbildung 19	Abbildung 20	Abbildung 21
		
<p>Triadisches <i>participation framework</i> vor Sequenzbeginn</p>	<p>Ver: es stand in den SPIELregeln;</p>	<p>Jhn: SO: oder wat; (1.0)</p>

Während John seine Frage formuliert, behält er sowohl die Körperposition als auch die Blickausrichtung bei. Die Adressatenauswahl kann aufgrund des dyadisch reorganisierten *participation framework* und des gemeinsamen Wissens um Verenas Expertenstatus ohne namentliche Adressierung oder turnzuteilenden Blick erfolgen. Anstelle einer visuellen Adressierung signalisiert der statisch auf den Golfball gerichtete Blick im Verbund mit der eingefrorenen Körperhaltung, dass der Sprecher von einem Monitoring seiner körperlichen Displays ausgeht und diese zum Zweck genauer visueller Wahrnehmung durch die Adressatin aufrecht erhält. Der Verweisraum für das Modaldeiktikon *so* ist in diesem Fall folglich der gesamte Körper des Sprechers als Display-Raum einer richtigen oder falschen Golfspielerhaltung. Die auf verbaler Ebene formulierte Frage setzt als zweiten Paarteil eine Antwort relevant, die nur und erst dann gegeben werden kann, wenn die durch das Modaldeiktikon erteilte perzeptorische Anweisung erfüllt worden ist. Anders als bei den oben betrachteten Instruktionssequenzen, in denen die Rollen von Sprecher, Zeigendem und Instruierendem in einer Person vereinigt waren, haben wir hier den Fall, dass der Sprecher zwar zugleich Zeigender ist, doch was er zeigt bzw. ausführt, ist potentiell korrekturbedürftig: Durch die multimodale Integration von Körper-Display, Frageintonation auf dem Deiktikon *so* und die Expansion durch *oder wat* fordert er seine Adressatin nicht allein zur Wahrnehmung seiner Haltung auf, sondern als weiter gespannte Projektionen zu einer Evaluation und ggf. zu entsprechenden Korrektur- und Instruk-

tionshandlungen. Die an die Adressatin gerichtete Aufforderung lautet, ihn wahrzunehmen und im Hinblick auf Fehlhaltungen, die nur ihrem Expertenblick zugänglich sind, zu korrigieren. Wie der weitere Verlauf deutlich macht, werden diese Projektionen interaktiv eingelöst.

Abbildung 21 zeigt, dass Verena unmittelbar nach Johns Turnabschlusspunkt ihre statische Standposition auflöst und in der einsekündigen Pause (Z. 9) zu ihm läuft. Simultan zu ihrer Artikulation des Modaldeiktikums (Z. 9: *SO:?*) tritt sie an John heran und beginnt, Oberkörper und Arme in eine Golfschläger haltende Position zu bringen (Abbildung 22). Mit dem Sprecherwechsel (Z. 9) werden die Rollen Sprecherin, Zeigende und Instruierende wieder vereint. Ein wichtiger Unterschied zu den in Abschnitt 3.3. dargestellten Instruktionshandlungen besteht darin, dass das Modaldeiktikon von der Instruierenden hier mit ansteigender *try marker*-Intonation (Auer 1979; Sacks/Schegloff 1979) gesprochen wird. Damit wird über die Wahrnehmungsaufforderung hinaus ein Verstehensdisplay des Adressaten relevant setzt. Dieses gedehnte, intonatorisch ansteigende *so* tut aber noch mehr, denn es projiziert darüber hinaus eine Demonstrationshandlung und zeigt dem Adressaten an, dass zu deren Ermöglichung bestimmte Schritte der interpersonellen Koordinierung bewerkstelligt werden müssen. D.h. dieses *so* kontextualisiert in diesem Fall keine simultan stattfindende körperliche Aktivität, sondern projiziert eine, die sich erst im Preparationsstadium befindet und deren Gelingenbedingungen gemeinsam hergestellt werden müssen. Dazu gehört die Übergabe des Schlägers von John an Verena in der anschließenden 0.5-sekündigen Pause, in der sich Verena außerdem parallel zu John ausrichtet und dabei die Oberkörper- und Armhaltung einer Golfschlagenden einnimmt (Abbildung 23). Es folgen Demonstrationen (Z. 11-15), in denen Verena John die richtige Schlägerhaltung vorführt. Die körperlichen Demonstrationen werden verbal durch verschiedene gestisch gebrauchte Deiktika begleitet (Z. 12: *SO:*, Z. 13: *den*, Z. 13: *HIERhin*, Z. 14: *SO*), die an dieser Stelle nicht im einzelnen analysiert werden können.

In der anschließenden 0.8-sekündigen Pause bringt Verena ihre Fuß- und Körperposition in ein *side-by-side arrangement* (Kendon 1990: 213) mit John, bevor sie zum Schlag ausholt. Diese Schlagbewegung wird verbal durch die Ankündigung *un' (-) und dann SO;* (Z. 16) projiziert. Beim Temporaldeiktikum *dann* verlässt der Schläger die Ruheposition und wird nach rechts seitlich hochgeschwungen (Abbildung 24). Die Amplitude erreicht mit dem Turnabschluss auf *SO* ihren Umkehrpunkt. Jedoch wird die anschließende, abwärts laufende Schlagbewegung nicht zu Ende geführt, sondern auf der Hälfte abgebremst und angehalten. Wie im Ananasbeispiel (Abschnitt 3.3) wird die relevante Handlung nicht wirklich ausgeführt, sondern zu didaktischen Anschauungszwecken lediglich simuliert.

Abbildung 22	Abbildung 23	Abbildung 24
		
Ver: SO:?	(0.5)	un' (-) und dann SO;

Das Besondere an dem Golf-Beispiel liegt darin, dass die multimodale Kollaboration zwischen den Beteiligten mit einem Minimum an Sprache vollzogen wird und in den Zeilen 7-16 quasi auf einen Dialog mit hochgradig kontextualisierungsreichen *so*-Verwendungen reduziert ist, die jeweils unterschiedliche Folgehandlungen projizieren. Die Projektionsverhältnisse

gewinnen dadurch an Komplexität, dass ihr Anforderungsprofil koordinations-theoretisch in höherem Maß nicht nur intraindividuell, sondern interindividuell organisiert ist. So kontextualisiert Johns Frage *SO: oder wat* (Z. 7) nicht nur sein körperliches Display als Golfspieler, sondern sie projiziert über die Anweisung zur visuellen Wahrnehmung hinaus eine Evaluation, Korrektur und Instruktion seitens der Adressatin. Die Antwort der Adressatin erfolgt nonverbal, indem sie sofort auf John zugeht und damit ihre Bereitschaft zur Übernahme der Instruktorrolle körperlich zum Ausdruck bringt. Ihr anschließend mit *try marker*-Intonation produziertes *SO:?* (Z. 9) fokussiert nun umgekehrt seine visuelle Wahrnehmung auf ihre körperlichen Ausdrucksformen und projiziert darüber hinaus erstens die zur Demonstration notwendigen Vorbereitungshandlungen (Schlägerübergabe), zweitens die Demonstration selbst und drittens ein Verstehensdisplay ihres Adressaten. Dabei sind die Verbindungsbögen zwischen den von Verenas *SO:?* aufgebauten Projektionen und den projektionsabarbeitenden Elementen unterschiedlich weit gespannt. Sie unterscheiden sich nicht nur hinsichtlich der Realisierungsmodalitäten (verbal – körperlich/visuell) und hinsichtlich ihres koordinations-theoretischen Status (interpersonell: visuelle Wahrnehmung, Schlägerübergabe, Verstehensdisplay; intrapersonell: körperliche Demonstration), sondern auch hinsichtlich ihrer Temporalität. Während die Schlägerübergabe als Voraussetzung für die körperliche Demonstration sofort stattfindet, erfolgt ein verbales Verstehensdisplay durch minimale Rückmeldesignale erst in Z. 21 (*hm=hm,*) und Z. 25 (*hm,*), nachdem Verena zunächst eine Rückversicherungsformel (Z. 19: *WEISSte*), dann eine Expansion ihrer Instruktion (Z. 22/23) und eine weitere Rückversicherungsformel in Form des tags *ne?* (Z. 24) produziert hat.

In dieser Sequenz handelt es sich beim Demonstrieren und Instruieren auch um ein Vormachen zum Zwecke des Nachmachens. D.h. die Beteiligten wechseln mehrfach ihre Rollen als Betrachtende/r und Performierende/r. Dies erfordert in weit höherem Maß als in dem Ananas- und dem Spaghetti-Beispiel (Abschnitt 3.3.) ein fortwährendes, intensives *wechselseitiges* Monitoring. In der *online*-Emergenz dieses mit multiplen Aktivitäts- und Rollenwechseln einhergehenden Prozesses, die von den Beteiligten gemeinsam hergestellt werden müssen, fungiert das deiktisch-gestisch verwendete *so* als aktivitätsstrukturierendes, wahrnehmungssteuerndes, rollenzuweisendes und die interpersonelle Koordination organisierendes, hochgradig kontextsensitives Instrument.

4. Schlussdiskussion

Meine Analyse erfolgte mit dem Ziel, die komplexen Projektionsverhältnisse zwischen verbalen und visuellen Ausdrucksressourcen beim deiktisch-gestischen Gebrauch des Ausdrucks *so* darzulegen. Dabei habe ich anhand meiner Daten vier verschiedene Konstruktionstypen ermittelt, die sich durch die Art und Weise unterscheiden, wie das Deiktikon *so* in den verbalen Kontext eingebettet ist. Die Darstellung begann mit dem verbal explizitesten Format (Abschnitt 3.1.), sie setzte sich fort mit zwei multimodal ganz ähnlich organisierten Formaten (Abschnitt 3.2. und 3.3.), die sich in der Hauptsache durch das *speech-affiliate* der ikonischen Geste voneinander unterscheiden (X = Adjektiv in Abschnitt 3.2.; X = Verb in Abschnitt 3.3.), und sie endete mit dem verbal am stärksten reduzierten Format (Abschnitt 3.4.).

Bei dem ersten von mir als *visuelle Evidenz*-Konstruktion bezeichneten Typ (*SO sieht/sehen X aus + Präsentationsgeste*) ist das multimodale Realisierungsformat vereinfachend zusammengefasst wie folgt organisiert: Blick auf Wahrnehmungsraum bzw. -objekt → gefolgt von *so* + simultaner Präsentationsgeste → gefolgt von VP → gefolgt von NP. Dabei wird in der Verbalphrase durch ein visuelles Wahrnehmungsverb (in der Regel *sehen*) die vom Adressaten auszuführende Perzeptionshandlung angewiesen. Die den Gegenstand X bezeichnende NP ist der Präsentativgeste sequentiell nachgeordnet. Die Besonderheit dieser Konstruktion liegt darin, dass mit ihr auf die Prototypikalität des gestisch präsentierten X fokussiert wird. Der Adressat wird nicht nur angewiesen, seine visuelle Aufmerksamkeit auf

den Gegenstand zu richten, sondern er soll seine Wahrnehmung dergestalt organisieren, dass er den Gegenstand in seiner Phänomenologie, seinen gestaltbildenden Merkmalen, wahrnehmen und als exemplarischen Vertreter der entsprechenden Kategorie kognitiv repräsentieren kann.

Anders verhält es sich beim zweiten Konstruktionstyp *SO X (X = Adjektiv) + Geste*. In diesem Fall soll der Gegenstand nicht holistisch wahrgenommen, sondern im Hinblick auf die durch die ikonische Geste repräsentierte und verbal durch das Adjektiv X bezeichnete Eigenschaft fokussiert werden. Hinsichtlich der temporalen Relationen konnte festgestellt werden, dass das semantische *speech-affiliate* – das Adjektiv – der entsprechenden ikonischen Geste nachfolgt, so dass sich verallgemeinernd folgendes Projektionsschema ergeben hat: Blick auf Wahrnehmungsraum/-objekt → projiziert Onset der Geste → ikonische Geste simultan zu *so* →→ beide projizieren auf unterschiedlichen Modalitätsebenen ein eigenschaftsbezeichnendes Adjektiv → Projektionseinlösung durch Adjektiv.

Mit dem zweiten Konstruktionstyp verbindet den dritten Konstruktionstyp *SO X (X = Verb) + Geste* erstens das multimodale Realisierungsformat (*speech affiliate* – hier das handlungsbezeichnende Verb – nach der Geste), und zweitens der spezifische Gestentyp der ikonischen Gesten. Die Ikonizität beruht in diesem Fall nicht auf einer Ähnlichkeitsrelation zwischen Gestenform und adjektivisch bezeichneter Eigenschaft wie beim zweiten Typ, sondern auf einer vollzugsperspektivischen Repräsentationsbeziehung zwischen einer körperlich ausgeführten oder simulierten Handlung und dem entsprechenden Handlungsverb. Dem Deiktikon *so* kommt dabei die Funktion zu, denjenigen Moment in der emergierenden körperlichen Demonstration, den es visuell wahrzunehmen gilt: also die eigentliche Vollzugsgeste selbst, exakt zu indizieren.

Als vierter Konstruktionstyp wurde der verbal isolierte Gebrauch von *SO + Geste/Performanz/Inszenierung* ermittelt. Der verbalen Reduziertheit steht ein hoher körperlich-visueller Realisierungsaufwand gegenüber, denn in der Regel wird durch das Deiktikon eine gesamtkörperliche Performanz des Sprechers projiziert, bei der dessen Körper einen *semiotischen Doppelraum* konstituiert. Er fungiert zugleich als Performanzsubjekt und als Performanzobjekt. Dabei sind die Übergänge zur Inszenierung fließend. So finden sich im Kontext narrativer Figurendarstellung in Alltagserzählungen Fälle einer multimodalen Inszenierung von Figuren, die über die Redewiedergabe hinausgehen, indem der Erzähler mit der animierten Rede zugleich die körperlichen Verhaltensweisen der erzählten Figur visuell zur Anschauung bringt (Stukenbrock i. Dr.). Bei diesen mit *so* eingeleiteten Inszenierungen, die hier nicht diskutiert werden konnten, sind bemerkenswerte Familienähnlichkeiten zur Quotativ-Konstruktion mit *so* (Auer 2006) festzustellen, die ausführlicher untersucht zu werden verdienen.

Verallgemeinernd ist festzuhalten, dass der deiktisch-gestische Gebrauch des Ausdrucks *so* weit mehr leistet als lediglich eine Perzeptionsanweisung an den Adressaten darzustellen. Je nachdem, in welchem verbalen und interaktiven Kontext es verankert ist, kommen noch andere, konstruktionsspezifische Projektionen mit ins Spiel. Bei der *visuellen Evidenz-Konstruktion* wird die Perzeptionsanweisung dahingehend spezifiziert, dass das präsentierte Objekt phänomenologisch und holistisch wahrgenommen und als Prototyp kognitiv repräsentiert bzw. in den Wissensbestand integriert werden soll. Bei der ad-adjektivischen Konstruktion geht es demgegenüber um die perzeptuelle Fokussierung auf ein einem individuellen Gegenstand akzidentell eignendes Merkmal. Die syntagmatische Kombination des Deiktikons *so* mit einem Verb wiederum kontextualisiert eine Vollzugsgeste oder -handlung, die mit der Vermittlung von prozessuellem Wissen insbesondere in Instruktionssequenzen verbunden ist. Das syntaktisch unverbundene *so* schließlich kontextualisiert eine gesamtkörperliche Performanz oder Inszenierung, die nicht nur wahrgenommen, sondern je nach Aktivitätstyp (z.B. Instruktion, Narration) evaluiert, korrigiert, applaudiert usw. werden soll. Wie an dem Golf-

Beispiel zu sehen war, spielt die Intonation als Kontextualisierungsmarker dabei eine Schlüsselrolle.

Allen vier Typen ist gemeinsam, dass das Deiktikon *so* als Kontextualisierungshinweis auf eine vom Sprecher ausgeführte und vom Adressaten visuell wahrzunehmende Handlung fungiert. Dies tut es jedoch nicht allein. Als interaktiv ebenso wichtige Herstellungshandlungen haben sich in allen Fällen die Konstituierung des Interaktionsraums, die Körperorientierung der Beteiligten sowie der Einsatz des Blicks als körperlich-visuelles Verweisinstrument erwiesen. Die Analysen belegen daher eine untrennbare Verbindung zwischen der *online*-Organisation des Interaktionsraums, des Wahrnehmungsraums, des Körper-, Blick- und Gesteneinsatzes und der Sprachverwendung. Das bedeutet, dass der Kontext keine präexistente, statische Größe darstellt, sondern als Resultat interaktiver Hervorbringungen (Hanks 1992: 70; Mondada 2007: 86) mittels unterschiedlicher Ressourcen zu begreifen ist. Er kann sich in der *online*-Emergenz der Interaktion ständig verändern und muss daher in dieser Veränderbarkeit begriffen werden. Dies erfordert eine Theorie der gesprochenen Sprache, die die zu seiner Konstitution eingesetzten körperlich-visuellen Ressourcen einer ebenso ernsthaften Analyse unterwirft wie die Sprache selbst und face-to-face-Kommunikation als komplexes Zusammenspiel von projektionsaufbauenden und projektionsabarbeitenden Interaktionselementen auf allen Ebenen modelliert. Nur wenn die Theorie alle diese Aspekte integrativ behandelt, kann sie der Konstitutionslogik multimodal organisierter face-to-face-Kommunikation gerecht werden.

5. Literaturverzeichnis

- Auer, Peter (1979): Referenzierungssequenzen in Konversationen: das Beispiel 'Ortsangaben'. In: *Linguistische Berichte* 62, 94–106.
- Auer, Peter (1981): Zur indexikalitätsmarkierenden Funktion der demonstrativen Artikelform in deutschen Konversationen. In: Hindelang, Götz/Zillig, Werner (Hgg.): *Sprache: Verstehen und Handeln*. Tübingen: Niemeyer, 301–310.
- Auer, Peter (1986): Kontextualisierung. In: *Studium Linguistik* 19, 22–47.
- Auer, Peter (2000): *On line*-Syntax - Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen. In: *Sprache und Literatur in Wissenschaft und Unterricht* 85, 43–56.
- Auer, Peter (2005): Projection in interaction and projection in grammar. In: *Text*, 25:1, 7–36.
- Auer, Peter (2006): Construction Grammar meets Conversation: Einige Überlegungen am Beispiel von "so"-Konstruktionen. In: Günthner, Susanne/Imo, Wolfgang (Hgg.): *Konstruktionen in der Interaktion*. Berlin, New York: de Gruyter, 291–314.
- Auer, Peter/Di Luzio, Aldo (Hgg.) (1992): *The Contextualization of Language*. Amsterdam, Philadelphia: Benjamins.
- Bühler, Karl (1934/1965): *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Stuttgart: Gustav Fischer Verlag.
- Deppermann, Arnulf/Schmitt, Reinhold (2007): Koordination. Zur Begründung eines neuen Forschungsgegenstandes. In: Schmitt, Reinhold (Hg.): *Koordination. Analysen zur multimodalen Interaktion*. Tübingen: Narr, 15–54.
- Ehlich, Konrad (2007): *Sprache und sprachliches Handeln*. Berlin, New York: de Gruyter.
- Eisenberg, Peter (2006): *Grundriss der deutschen Grammatik. Bd.2: Der Satz*. Stuttgart: Metzler.

- Fillmore, Charles F. (1997): *Lectures on Deixis*. Stanford: CSLI Publications.
- Fricke, Ellen (2007): *Origo, Geste und Raum. Lokaldeixis im Deutschen*. Berlin, New York: de Gruyter .
- Glück, Helmut (Hg.) (2005): *Metzler-Lexikon Sprache*. 3., neubearb. Aufl. Stuttgart, Weimar: Metzler.
- Goffman, Erving (1981): *Forms of talk*. Philadelphia: University of Pennsylvania Press.
- Goodwin, Charles (2003): Pointing as Situated Practice. In: Kita, Sotaro (Hg.): *Pointing: Where language, culture, and cognition meet*. Mahwah, N.J.: Erlbaum, 217–241.
- Gülich, Elisabeth/Kotschi, Thomas (1987): Reformulierungshandlungen als Mittel der Textkonstitution. Untersuchungen zu französischen Texten aus mündlicher Kommunikation. In: Motsch, Wolfgang (Hg.): *Satz, Text, sprachliche Handlung*. Berlin: Akademie-Verlag, 199–361.
- Hanks, William (1992): The Indexical Ground of Deictic Reference. In: Duranti, Alessandro/Goodwin, Charles (Hgg.): *Rethinking Context. Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press , 43–76.
- Harweg, Roland (1990): *Studien zur Deixis*. Bochum: Brockmeyer.
- Heath, Christian (1986): *Body movement and speech in medical interaction*. Cambridge: Cambridge University Press.
- Hennig, Mathilde (2006): So, und so, und so weiter. Vom Sinn und Unsinn der Wortklassifikation. In: *Zeitschrift für germanistische Linguistik* 34, 409–431.
- Kendon, Adam (1990): Spatial organization in social encounters: the F-formation system. In: Kendon, Adam (Hg.): *Conducting interaction. Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press, 209–237.
- Kendon, Adam (2004): *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, Adam/Versante, Laura (2003): Pointing by Hand in "Neapolitan". In: Kita, Sotaro (Hg.): *Pointing: Where language, culture, and cognition meet*. Mahwah, N.J.: Erlbaum, 109–137.
- Kita, Sotaro (Hg.) (2003): *Pointing: Where language, culture, and cognition meet*. Mahwah, N.J.: Erlbaum.
- Lyons, John (1977): *Semantics*. Cambridge: Cambridge University Press.
- McNeill, David (1992): *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, David (2005): *Gesture and Thought*. Bristol: University of Chicago Press.
- Mondada, Lorenza (2006): Participants' online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. In: *Discourse Studies* 8:1, 117–129.
- Mondada, Lorenza (2007): Interaktionsraum und Koordinierung. In: Schmitt, Reinhold (Hg.): *Koordination. Analysen zur multimodalen Interaktion*. Tübingen: Narr, 55–93.
- Sacks, Harvey/Schegloff, Emanuel A. (1979): Two Preferences in the Organization of Reference to Persons in Conversation and Their Interaction. In: Psathas, George (Hg.): *Everyday language. Studies in Ethnomethodology*. New York: Irvington, 15–21.
- Sandig, Barbara (1987): Kontextualisierungshinweise: Verwendung von so im Prozeß sprachlichen Handelns. In: Rosengren, Inger (Hg.): *Sprache und Pragmatik*. Stockholm: Almq-

vist & Wiksell, 327–333.

Schegloff, Emanuel A. (1984): On some gestures' relation to talk. In: Atkinson, John Maxwell/Heritage, John (Hgg.): *Structures of Social Action. Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 266–296.

Schmitt, Reinhold (Hg.) (2007): *Koordination. Analysen zur multimodalen Interaktion*. Tübingen: Narr.

Schmitt, Reinhold/Deppermann, Arnulf (2007): Monitoring und Koordination als Voraussetzungen der multimodalen Konstitution von Interaktionsräumen. In: Schmitt, Reinhold (Hg.): *Koordination. Analysen zur multimodalen Interaktion*. Tübingen: Narr, 95–128.

Sennholz, Klaus (1985): *Grundzüge der Deixis*. Bochum: Brockmeyer.

Streeck, Jürgen (1995): On projection. In: Goody, Esther N. (Hg.): *Social intelligence and interaction. Expressions and implications of the social bias in human intelligence*. Cambridge: Cambridge University Press, 87–110.

Streeck, Jürgen (2002): Grammars, Words, and Embodied Meanings: On the Uses and Evolutions of So and Like. In: *Journal of Communication* 52, 581-591.

Stukenbrock, Anja (i. Dr.): Zur Beredsamkeit des Körpers. Figurendarstellung und Figurenwissen als multimodale Alltagsinszenierung. In: Lampert, Fabian et al. (Hg.): *Figurenwissen. Funktionalisierung und Repräsentation von Wissen bei der narrativen Figurendarstellung*. Berlin, New York: de Gruyter .

Stukenbrock, Anja (2008): "Wo ist der Hauptschmerz?" - Zeigen am eigenen Körper in der medizinischen Kommunikation. In: *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 9, 1–33.

Stukenbrock, Anja (2009): Referenz durch Zeigen. Zur Theorie der Deixis. In: *Deutsche Sprache* 37, 289–316.

Thurmair, Maria (2001): *Vergleiche und Vergleichen. Eine Studie zu Form und Funktion der Vergleichsstrukturen im Deutschen*. Tübingen: Niemeyer.

Weinrich, Harald (1993): *Textgrammatik der deutschen Sprache*. Mannheim, Leipzig, Wien, Zürich: Duden Verlag.

Zifonun, Gisela et al. (1997): [= IdS-Grammatik] *Grammatik der deutschen Sprache*. Berlin, New York: de Gruyter.