

## #5. Secondary articulation and sound change

chang01.pdf

+++++

Es geht hauptsächlich darum, zu prüfen, weshalb einige Verwechslungen, die zu Lautwandel führen, asymmetrisch sind. Z.B. findet man vor hohem Vokal häufig k -> t, aber selten t -> k. Der vermutliche Grund dafür ist, dass /t/ und /k/ vor hohem Vokalen akustisch recht ähnlich sind, abgesehen von dem 'mid-frequency peak burst' der für /k/ charakteristisch ist.

Wichtig auf S. 80 die Analogie zur visuellen Wahrnehmung: "E" wird eher mit "F" verwechselt als umgekehrt (weil "E" die Merkmale von "F" mit einschließt).

Es gibt ähnliche Asymmetrien in der gesprochenen Sprache (vor allem /ki/ (wie E) -> /ti/ (wie F)).

Die Autoren wollen auch zeigen, dass die Grundlage dieser Verwechslung/dieses Lautwandels nichts mit der sogenannten linguistischen Markiertheit ('markedness') zu tun hat, d.h. markedness ist keine Erklärung, sondern höchstens eine Beschreibung (und dann auch nicht eine besonders gute) für solche Lautwandel. Hier einiges mehr zu markedness, solltet Ihr mehr dazu wissen wollen:

<http://grammar.about.com/od/mo/g/Markedness.htm>

Im Paper wollen Chang et al auch erklären, wieso der Lautwandel meistens /ki/ -> /tʃi/ ist, während man dagegen empirisch im Labor meistens /ki/ -> /ti/ feststellt.

Expt 1.

-----

Ein Merkmal von /k/ ist der mid-frequency spectral peak. Dieser wurde von /k/ entfernt und dem /t/ hinzugefügt. Die experimentellen Details zu dem Filter usw. sind für uns nicht wichtig. Am wichtigsten sind die Ergebnisse: /ki/ ohne mid-frequency peak wird manchmal als /ti/ wahrgenommen, obwohl /ki/ oft korrekt identifiziert wird, da andere Merkmale (wie eine lange VOT) für die Identifizierung von /k/ vorhanden sind. /ti/ mit dem Peak wird aber selten mit /ki/ (oder genauso oft mit /pi/) verwechselt.

Expt. 2

-----

Hier geht es darum, nachzuweisen, dass diese asymmetrische Verwechslung eher vor hohem vorderen Vokalen vorkommt. Die gefilterten /ki/ Stimuli wurden überwiegend als alveolar erkannt; einige als /tʃ/ (eventuell diejenigen mit einer langen VOT).

Expt. 3

-----

Hier wird versucht, nachzuweisen, dass der Lautwandel /ki/ -> /tʃi/ aufgrund der langen Aspirationsphase in /k/ vorkommt. Wenn dies der Fall ist, dann müsste es keine /gi/ -> /dʒi/ Verwechslung geben, da /g/ unaspiriert ist. Dies war auch der Fall.

ohala77

+++++

Viele Phonologen akzeptieren nicht, dass [w] labial-velar ist, weil das nicht sehr gut zur phonologischen Theorie passt, d.h. für viele Phonologen ist [w] entweder labial oder velar (oder velar mit sekundärer Artikulation).

Einige Phonologen (wie Andersen) behaupten z.B., dass /kp/ als labial fungiert, wenn im Inventar /p/ nicht vorhanden ist.

Die Analyse von 55 Sprachen, die in ohala77 durchgeführt wird, zeigt jedoch, dass es kaum Evidenzen dafür gibt. Die Details der Tabellen können überflogen werden. Das wichtigste ist wie folgt: Phonologen behaupten, dass /kp/ hauptsächlich dann vorkommt, wenn /p/ in einem Lautinventar fehlt. Wenn dies so wäre, müsste es mehr Sprachen mit /kp/ und ohne /p/ geben, als durch die Kombination der Wahrscheinlichkeiten vorhergesagt wird. d.h. wenn die Wahrscheinlichkeit, dass kein /p/ vorkommt  $p_1$  ist, und die Wahrscheinlichkeit, dass /kp/ vorkommt  $p_2$  ist, dann müsste die Wahrscheinlichkeit, dass eine Sprache kein /p/ UND /kp/ hat, also  $p(\text{kein /p/ UND /kp/})$ , größer als  $p_1 \times p_2$  sein. Dies ist nicht der Fall (Table I:  $p(\text{kein /p/}) = .2$ ;  $p(\text{/kp/}) = .69$ , daher  $p(\text{kein /p/ UND /kp/}) = .2 * .69 = .138$ , sodass  $0.138 * 55$  Sprachen heißt: zwischen 7 und 8 Sprachen müssten ein solches Muster haben; in der Tat sind es nur 3. Daher gibt es keine Evidenzen, dass /kp/ vorkommt, wenn es kein /p/ gibt.

Der Rest vom Paper behandelt die akustischen Gründe, weshalb /w/ sich manchmal als velar im Kontext von Nasalen zeigt, jedoch als labial vor Frikativen. Der erste Teil (A) versucht zu erklären, wieso doppelte Artikulationen so oft labial-velar sind - weil die akustischen Wirkungen von velar und labial sich gegenseitig verstärken: 'labiovelars are so popular: they push the second formant towards an extreme low value and thus produce sounds which are auditorily very distinct from other speech segments.'

Dann Generalisierung (B): Nasale werden meist zu Velaren vor [w]. Das ist so, weil [w, ŋ] akustisch sehr viele Ähnlichkeiten haben (in beiden Fällen gibt es kaum einen Einfluss von Anti-Resonanzen wegen des Oraltrakts)

Generalisierung (C): Frikative werden meistens zu Labialen vor [w]. Das ist so, weil, wenn ein Frikativ in der Umgebung von [w] erscheint, der Einfluss auf die Wahrnehmung vom labialen stärker ist als jener vom velaren Geräusch. Das wichtigste ist 4. auf S. 588. Das velare Geräusch ist kaum wahrnehmbar, weil es hinter der labialen Verengung produziert wird.

Es gibt weitere Beweise aus dem Bereich des Lautwandels, dass die akustische Wirkung der velaren Verengung nicht so prominent ist wie jene der labialen.

C: (S. 589): labial-velare Plosive entwickeln sich meistens zu Labialen.

D: vor Vokalen bleibt die Labialisierung, aber die Verengung im Vokaltrakt kann sehr variabel sein (variiert mit der Position der Zunge des Vokals).

Der Text 'Now we need to introduce a few qualifications...' bis zu Conclusions kann überflogen werden.

## Schlussfolgerung

-----

Die Tatsache, dass labial-velare Laute manchmal als labiale oder manchmal als velare erscheinen, entsteht nicht aufgrund von Mustern (markedness) in der Phonologie, sondern ist physiologisch-akustisch bedingt.

ohala06

+++++++

Vieles von diesem Paper kann überflogen/weglassen werden. Das wichtigste:

1. siehe S. 12 u. 13 für zusätzliche Beispiele, die für ohala77 relevant sind.
2. Phonetische Natürlichkeit = wieso es Präferenzen für gewisse Laute in den Sprachen der Welt gibt. Z.B. wie oben: wieso labial-velare sich als velar im Bereich von Nasalen, jedoch als labial in der Umgebung von Frikativen zeigen.

Wie in ohala77 und auch in diesem Paper argumentiert, kommen solche Präferenzen (Natürlichkeit) wegen physischer Prinzipien der Akustik und Sprachproduktion zustande. Die Phonologie hingegen versucht die Natürlichkeit aufgrund von Regeln zu erklären. Vor allem ist es ein Prinzip der Generativen Phonologie, vorhersagbare phonetische Aspekte zu entfernen. Da Obstruenten (Plosive und Frikative) = Nicht-Sonoranten = [-son] so oft stimmlos sind, ist die Spezifizierung redundant, ob sie stimmlos oder stimmhaft im Lexikon eingetragen werden. Das Stimmhaftigkeitsmerkmal im Lexikon ist daher einfach [voice] und es gibt eine Regel, die die phonetische Details hinzufügt (dass [voice] zu [-voice] wird, wenn der Konsonant [-son] ist).

Ohalas Kritik:

1. das ist eine Beschreibung und widerspiegelt nicht die zugrunde liegenden physischen Prinzipien.
2. es gibt kaum psychologische Evidenzen, dass z.B. das Lexikon und die kognitiv gespeicherten Kenntnisse der Sprache minimal (ohne phonetische Details) sind.
3. nicht alle physischen Prinzipien müssen kognitiv sein. z.B. wir essen und verdauen, ohne dass Verdauung als Regeln kognitiv erfasst werden muss. Ein Fels unterliegt den Prinzipien der Physik (ebenso wie die Sprachproduktion und -perzeption), hat aber keine Kognition!
4. Die Bausteine (primitives) um phonetische Natürlichkeit zu erklären, sind nicht abstrakte Merkmale wie [-son] oder [+cont], sondern physische Prinzipien. Daher kann z.B. eine Regel (S. 11) nicht erklären, wieso ein Obstruent [-cont] zwischen zwei [+cont] Lauten auftaucht, wie z.B. [ls] -> [lts] oder [sl] -> [stl].

## #6. Perception, syllable-structure, and sound change

ohala90.pdf

+++++

Es handelt sich um die Beziehung zwischen Perzeption, Assimilation, und Lautwandel. Assimilation zB. "wann" -> /wan/ in "wann kann ich"...

Beispiele der Beziehung zwischen Assimilation und Lautwandel auf S 258.

Lautwandel, die mit Assimilationen verbunden sind, sind meistens von rechts nach links (d.h. der linke K passt sich bezüglich der Artikulationsstelle dem rechten an).

Die Hypothese ist, dass diese Beziehung zwischen Assimilation und Lautwandel sich perzeptiv erklären lässt.

Es gibt mehrere Beweise, dass in VC-CV Sequenzen (C ist ein Plosiv), die Cues zu C in CV perzeptiv wesentlich dominanter sind als diejenigen in VC. Einige Gründe dafür – S. 261. Expt. 1. VC wurde mit CV zusammengesplacet für verschiedene Kombinationen von Artikulationsstellen in Plosiven (zB /atka/) und in Nasal-Plosiv-Reihenfolgen (/anpa/ usw) - S. 263.

Hörer mussten entscheiden, welche Artikulationsstelle sie hörten. D.h. bei der Präsentation von zB /apka/ hatten sie die Wahl zwischen /apa/ oder /aka/ oder etwas anders. Die Ergebnisse zeigten, dass sich die Antworten überwiegend nach C2 richteten. Expt. 2: trägt nichts neues bei eigentlich, und kann überflogen/weggelassen werden. Der Rest von diesem sehr interessanten Artikel kann komprimiert werden ca. wie folgt:

Weder Markedness noch 'ease of articulation' können erklären, weshalb Assimilationen meistens in eine Richtung gehen. Das wichtige in diesem Paper ist: welche phonetischen/perzeptiven Gründe tragen dazu bei, dass Assimilation häufiger antizipatorisch (von rechts nach links) als regressiv sind? Dann gibt es das übliche von Ohala, nämlich inwiefern die formalen Regeln aus der Phonologie nicht imstande sind, die Beziehungen zwischen diachronen und synchronen Vorgängen in der sprachlichen Kommunikation zu erklären (sondern nur zu beschreiben).

ohala84b.pdf

+++++

es geht hier hauptsächlich darum, wieso CV salienter ist im Vergleich zu VC - zum Teil weil der Vokal und der Konsonant mehr miteinander überlappen in einer VC als CV-Reihenfolge. Verschiedene Gründe werden dafür angeführt, weshalb CV-Reihenfolgen perzeptiv salienter als VC-Reihenfolgen sind.

S. 113. Schönes Beispiel um die Unterschiede zwischen Phonologie und Phonetik zu erklären!

S. 115 Syllable Onset vs Coda ist zusätzlich nützliches Material für ohala90.pdf oben – weshalb Cs in prosodisch initialer Position perzeptiv dominanter sind als in finaler Position.

Das P-Center (geht übrigens auf Morton, Markus & Frankish, 1976, Psychological Review, 83, 405-408 zurück) ist der Zeitpunkt der Synchronisierung in der Metronomsprache (d.h. wo der Taktschlag vorkommt, wenn man eine Silbe pro Metronomschlag spricht). Der P-Center kommt nicht am akustischen Onset der Silbe vor, sondern befindet sich meistens irgendwo in den Onset-Cluster-Konsonanten vor dem Vokal. Für Carol Fowler ist übrigens das P-Center der \*artikulatorische\* Vokalansatz, der vor dem akustischen Onset des Vokals in dem K-Cluster auftritt.

Abs. 4 (Phonetic processes in the metrical grid) ist interessant, aber kann eventuell überflogen werden - es geht hier darum, eine Verbindung zwischen prosodischen Bäumen und kompensatorischer Kürzung zu erstellen (kompensatorische Kürzung: je mehr Silben im Wort oder Wörter in einer Phrase vorkommen, umso kürzer ist die Segmentdauer, vor allem von Vokalen: daher ist der Vokal kürzer in 'lengthier' als in 'lengthy' als in 'length').

Abs. 5 befasst sich mit dem Sonoritätsprinzip: dass die Sonorität gegen den Silbenkern (Vokal) zu- und dann abnimmt. Ohala argumentiert, dass das Sonoritätsprinzip wie in 5. auf S. 122 dargestellt keine Erklärung, sondern nur eine Beschreibung ist (siehe auch ohala92.pdf unten)

hura92.ls.pdf

+++++

Das Paper versucht zu erklären, weshalb Frikative (im Vergleich zu Nasalen) kaum assimilieren. Vielleicht hier vor allem die Assimilationssachen für deutsch auf S. 62 zusammenfassen. Im Experiment wird festgestellt, dass Fehler in der Identifizierung von Nasalen und Plosiven, kaum jedoch von Frikativen vor einem Konsonanten auftreten - und dies wird mit der Nicht-Assimilation von Frikativen verbunden: d.h. es geht um *perzeptive Salienz*.

Das Experiment war wie folgt. Logatom-Namen wurden produziert der Form Shani/m, n, ŋ, p, t, k, f, s, ʃ] [p, t, k erry]

Die Zweitnamen Perry, Terry, Kerry wurden auch nach Anna produziert, und initiale /p, t, k/s zu haben, die vom davorstehenden K unbeeinflusst waren.

Der erste Name wurde paarweise mit dem ( nach Anna produzierten) zweiten Namen kombiniert. Die Hörer mussten den ersten Namen identifizieren.

Die Ergebnisse: Frikative wurden meistens korrekt - oder zumindest zuverlässiger als Nasale und Plosive - identifiziert.

Die Sache mit Assimilation und Ohala (An assimilation of the pattern of confusions... S. 66) ist etwas verwirrend und nicht klar; und außerdem argumentieren die Autoren, dass die Ergebnisse dann irgendwie mit Ohalas Theorie konsistent sind. Daher vielleicht das weglassen.

In der Diskussion versuchen sie dann zu antworten, weshalb Frikative resistent gegen Assimilation sind. S. 68: es gibt keine Erklärung dafür in der Phonologie, hauptsächlich weil feature geometry nicht erklären kann, weshalb Nasale für Assimilation als Frikative anfälliger sind.

Die wichtige Zeile: "Our results show a distinct perceptual difference between nasal and stop consonants and fricatives - differences which parallel the likelihood of each class to undergo assimilation to a following stop".

d.h. Frikative sind perzeptiv salienter und werden aus diesem Grund von den Vorgängen der Assimilation kaum beeinflusst. Dies betrifft aber im wesentlichen Sibilanten /s, ʃ/ (S 69): Nicht-sibilanten wie /f, θ/ könnten eventuell genauso instabil wie Nasale sein, gerade weil sie perzeptiv salient sind.

Der Teil (S. 69-70) 'A perceptually based prediction kann überflogen/weggelassen werden.

ohala92.pdf

+++++

Aufbauend auf ohala84b.pdf: ein Angriff auf die Sonority hierarchy. Hier soll auch (s. 345) erklärt werden (bevor man von der Sonority hierarchy spricht) was phonotactics sind (wie definiert man phonotactics) und welche phonotactics von der Sonority hierarchy 'vernachlässigt' worden sind. Vielleicht dann erklären, was in diesem Kontext 'modulation' (S. 350) bedeutet – d.h. hier handelt es sich um ein Experiment, um zu zeigen, dass CCV Reihenfolgen mit der größten akustischen Modulation im F1 x F2 x F3 Raum gerade diejenigen sind, die in den Sprachen der Welt bevorzugt werden. 5.5 präsentiert Argumente, ob überhaupt die Silbe notwendig ist. Hier ist der Standpunkt, dass die Silbe nicht (wie in der Phonologie) ein gegebenes Konstrukt ist, sondern aus den Vokalktrakt-Schließungen/Öffnungen entsteht. Vielleicht dieses 5.5 überfliegen. 5.6 ist dann zusätzliches Material, das für die anderen drei Papers oben relevant ist.

## # 7. Automatic and planned aspects of speech

sole2007

-----

Es geht darum, inwiefern man zwischen mechanistischen, nicht geplanten Effekten der Sprachproduktion auf der einen Seite von geplanten Vorgängen auf der anderen trennen kann.

S. 303

intrinsic vs. extrinsic allophones

intrinsic sind diejenigen, die aufgrund der zeitlichen Überlappung der Vokalorgane zustande kommen (z.B. Lippenrundung in /t/ und Nasalisierung in /u/ in /tun/).

Extrinsische sind phonetische Varianten, die nicht dadurch erklärt werden können - z.B. Aspiration in deutsch, oder das velarisierte (dunkle) /l/ [ɫ] in English feel, steal usw.

S. 303-304. A physiologically based difference such as intrinsic differences in pitch or duration related to vowel height (with low vowels having a lower pitch and a longer duration than corresponding high vowels) may become part of a perceptual cue distinguishing different vowel heights.

Das bezieht sich auf die Phonologisierung: ein Effekt, der wegen Koartikulation entsteht, wird kontrastiv.

Vokalnasalisierung S. 307

-----

Es wird für Amerikanisch-Englisch vermutet, dass die Nasalisierung phonologisch ist - also geplanter Bestandteil des Vokals. d.h. ähnlich wie man planen muss, das Zungendorsum in einem /i/ in Richtung weichen Gaumen anzuheben, muss man auch planen, im Vokal das Velum zu senken, wenn Nasalisierung geplanter Bestandteil des Vokals ist. Wenn daher die Vokaldauer wegen Sprechgeschwindigkeitsänderungen kürzer/länger wird, dann müsste auch die Dauer der Nasalisierung im Vokal kürzer/länger werden, wenn Nasalisierung geplanter Bestandteil des Vokals im Amerikanisch-Englischen ist.

In vielen anderen Sprachen wie Spanisch ist Vokalnasalisierung eher mechanistisch, ca. wie folgt: damit die Velumöffnung im danach kommenden Nasalkonsonanten ein Maximum erreicht, muss die Velumsenkung schon im Vokal beginnen. Das ist aber nicht ein geplanter Bestandteil (eine Eigenschaft) des Vokals - daher gilt: wenn sich die Vokaldauer wegen Sprechgeschwindigkeit ändert, wird das den Zeitbeginn der Velumsenkung kaum beeinflussen.

All das wird mehr oder weniger mit Abb. 18.1 belegt. Die Velumsenkung beginnt früher im AmEngl. (oder Vokalnasalisierung ist länger) je schneller die Sprechgeschwindigkeit, aber nicht in Spanisch.

Stop Voicing S. 310

-----

Ähnliche Argumente wie für Vokalnasalisierung. Im Englischen ist Aspiration (und damit VOT) entscheidend für den Unterschied zwischen /p, t, k/ vs /b, d, g/ jedoch nicht im

Katalansischen - hier kann es einen positiven VOT im stimmlosen Plosiven geben, aber dies entsteht mechanistisch wegen des hohen oralen Luftdrucks vor der Verschlusslösung im katalansichen /p, t, k/. Daher müssten Änderungen in der Sprechgeschwindigkeit eher einen Einfluss auf englische, jedoch nicht katalanische VOT-Werte haben (weil Aspiration ein geplanter Bestandteil/eine Eigenschaft vom englischen (aber nicht vom spanischen)/p t k/ ist)/ Dies wird auch durch Abb 18.2 belegt.

Ebenfalls (S. 312) ändert sich die negative VOT als cue für /b, d, g/ in Sprachen wie Katalanisch mit der Sprechgeschwindigkeit, weil die negative VOT entscheidend ist, um in diesen Sprachen /b, d, g/ von /p, t, k/ zu differenzieren (also: negative VOT ist Bestandteil/Eigenschaft von /b, d, g/).

Vokaldauer als Cue zu postvokalischer Stimmhaftigkeit

-----

Phonetisch ist Vokaldauer kürzer vor stimmlosen im Vgl. zu stimmhaften Cs in VC Reimen (also wenn V und C in derselben Silbe ist). Dies ist aus dem folgenden Grund mechanistisch: der Verschluss ist in stimmlosen Cs länger und schneidet nach links den Vokal früher ab.

Im Englischen und teilweise im Deutschen (zB leiten vs. leiden) ist dieser Effekt phonologisiert worden, d.h. er hat sich zu einem entscheidenden (geplanten) Merkmal entwickelt, um post-vokalisch /p t k/ von /b d g/ zu differenzieren. Der Cue ist dann auch verstärkt (enhanced) worden: d.h. der Vokal in englisch 'sat' ist viel kürzer als man erwarten würde aufgrund solcher rein mechanistischen Effekte.

Also um es noch einmal klarer darzustellen.

Vokale sind aus mechanistischen (nicht-geplanten) Gründen in vielen Sprachen kürzer vor stimmlosen Cs. Dieser nicht-geplante Effekt ist im Englischen und Deutschen phonologisiert worden, d.h. Bestandteil des Vokals geworden. Eventuell als Folge der Phonologisierung wird der Vokal noch kürzer, damit der Unterschied zwischen sat/sad oder leiten/leiden verdeutlicht wird.

Solé erwähnt zusätzlich, dass der *proportionale* Vokaldauerunterschied vor stimmlosen und stimmhaften Ks konstant bleibt, *wenn Vokaldauer als Cue geplant wird*.

Z.B. nehmen wir an, dass die Vokaldauer in 'sat' 1.5 Mal größer ist als in 'sad'. Bei einer tatsächlichen Vokaldauer von 100 ms im betonten 'sat' müsste dann die Vokaldauer im betonten 'sad' 150 ms sein. Bei einer tatsächlichen Vokaldauer von nur 20 ms in unbetontem 'sat', ist die Vokaldauer in unbetontem 'sad' 35 ms, also ein Unterschied von nur 15 ms. D.h., der tatsächliche Unterschied in der Vokaldauer vor /t, d/ ist viel kleiner in unbetont vs. betont, *wenn Vokaldauer geplant ist*.

---

Diese Sache zum proportionalen Dauerunterschied wird sehr gut im nächsten Paper zusammengefasst (S. 608) 'For example, Caisse (1988), using the relative vowel duration as an index, found that the vowel duration differences before voiced and voiceless obstruents in American English could be modeled as a multiplicative effect [=

proportional], that is, the differences varied by a multiple of the duration of the vowel. In contrast, the duration of vowels before fricatives as opposed to stops showed an additive effect, that is, they differed by a constant number of ms. She suggested that multiplicative effects are centrally controlled [=geplant] and occur prior to additive effects [= mechanistisch], which are peripheral.'

---

Dies wird bestätigt durch die Abbildung auf S. 316. Der Unterschied zwischen voiceless/voiced ist viel größer in stressed als in unstressed - was darauf hindeutet, dass es sich um proportionale Vokaldauerunterschiede handelt (= 'sat' hat z.B. eine Dauer, die 1.5 Mal so gross ist wie bei 'sad'), was ebenfalls darauf hindeutet, dass Vokaldauer geplant ist.

Wenn hingegen Vokaldauer *nicht* geplant ist, dann sind die Unterschiede eher additiv. d.h. 'sat'/'sad' würden sich immer um z.B. 50 ms unterscheiden, unabhängig von den Daueränderungen, die wegen der Betonung verursacht werden. Das ist eindeutig für 'sat'/'sad' nicht der Fall, wie man in der Abbildung auf S. 316 sieht - aber es ist so für die Vokaldauerunterschiede vor Frikativen und Plosiven. D.h., die Vokaldauer ist etwas kürzer vor Plosiven als vor Frikativen. Dies ist aber mechanistisch/nicht geplant und daher additiv. Daher wird die Vokaldauer um ca. denselben Dauerwert von betont auf unbetont reduziert - und daher sind die gestrichelten und festen Linien unten in der Abbildung bei Frikativ und Plosiv parallel zu einander (während sie bei 'sat'/'sad' auseinander gehen).

Abb. 18.4: Hier wird wieder das Argument gebracht, dass Vokaldauer als Cue für die postvokalische Stimmhaftigkeit im Englischen jedoch nicht im Katalansichen geplant ist. Daher müsste der Unterschied in den Vokaldauern zwischen stimmlos und stimmhaft je nach Sprechgeschwindigkeit proportional im Englischen, jedoch additiv im Katalanischen variieren. Das zeigt auch Abb 18.4.

Das reicht dann auch für dieses Paper!

sole2010.pdf

=====

In diesem Artikel muss nur das wichtigste besprochen werden, d.h. all die statistischen Details usw. sollen überflogen werden/unerwähnt bleiben. Hier das meines Erachtens wichtigste:

In der Phonologie gibt es primäre diskontinutive Merkmale wie  $\pm$ nasal, aber auch sekundäre Merkmale, die aus den primären vorhersagbar sind (daher sind sie laut einiger phonologischen Theorien redundant). Die Frage ist: inwiefern werden diese sekundären Merkmale geplant?

## 2. Covarying features

Das wichtigste hier:

There are language-specific ways of dealing with the constraints imposed by the speech production system that have to be learned and controlled by the speaker.

d.h. es könnte sein, dass alle Merkmale geplant (also kognitiv) sind, weil sie immer von Sprache zu Sprache variieren (sie können also nicht nur als mechanistische Folge der Sprachproduktion entstehen, sonst würden sie nicht von Sprache zu Sprache variieren).

Es geht hier um den Einfluss der Vokalhöhe auf die Vokaldauer. Offenere Vokale haben eine längere Dauer, weil der Kiefer einen größeren Weg zurücklegen muss. Die Frage ist, ob solche Unterschiede geplant sind, oder ob sie eine mechanistische Folge der Sprachproduktion sind.

Die Hypothese:

wenn intrinsische Vokaldauer geplant ist, dann müsste die *proportionale* Dauer zwischen hohen, mittleren, und tiefen Vokalen relativ konstant bleiben.

D.h., wir erwarten eine statistische Interaktion zwischen z.B. Vokalhöhe und Geschwindigkeit, wenn Vokalhöhe geplant ist. Und zwar aus folgendem Grund:

Nehmen wir an, das Verhältnis hoch : mittel : tief = 2 : 1 : 0.5

Bei einer langsamen Geschwindigkeit könnten dann die Unterschiede so aussehen: 200, 100, 50 ms für hoch, mittel, tief.

Bei einer schnellen Geschwindigkeit vielleicht so: 100, 50, 25 für hoch mittel tief.

Daher ist z.B. der absolute Unterschied zwischen hoch und tief  $200 - 50 = 150$  ms in langsam, aber nur  $100 - 25 = 75$  ms in schnell.

Das würde in der Statistik (wenn Sie daran zurückdenken) zu einer Interaktion zwischen Vokalhöhe und Sprechgeschwindigkeit führen, weil die Unterschiede (in Millisekunden) anders in der langsamen als in der schnellen Sprechgeschwindigkeit sind.

Daher (!): Eine Interaktion in der Statistik zwischen einem Faktor (Vokalhöhe oder Vokallänge) und Sprechgeschwindigkeit deutet darauf hin, dass der Faktor *geplant* ist. Wenn keine Interaktion vorliegt, dann eher *nicht geplant*.

Daher, S. 620: 'if the long and short distinction is specified but not vowel height, one would expect interactions between rate and vowel length, but not between rate and vowel height.' [specified = geplant].

Die Autoren finden wie erwartet eine Interaktion zwischen Vokallänge (was kontrastiv ist im Japanischen) und Sprechgeschwindigkeit (S. 622 oben).

Die Umstand, dass 'the interaction between vowel height and speaking rate did not reach significance for any of the speakers' (S. 622), deutet also darauf hin, dass die Unterschiede in der Vokalhöhe *nicht geplant* sind. Daher S. 622/623: 'The fact that intrinsic durational differences remain relatively constant across changes in rate suggests that these are mechanical differences ...'

Man sieht auch die Interaktion in dem Bild Fig.1. Wenn keine Interaktion vorliegen würde, dann wären die 3 oberen Linien (ich nehme an diese sind für /a:, e:, i:/) parallel zu den unteren drei Linien verlaufen, und dies ist eindeutig nicht der Fall.

Den Rest bist zu 5.2.2 Catalan überfliegen.

Für Katalanisch:

Abb 2. Ich nehme an, dass die Linien von unten nach oben für: /i, e, ε, a/ stehen.

Hier sieht man also keine Interaktion zwischen /i, e/ (diese Linien sind zueinander parallel) und auch keine zwischen /ε, a/ (Linien auch parallel); aber es kommt doch zu einer Interaktion zwischen /i, e/ vs /ε, a/ (weil die /i, e/ Linien und die /ε, a/ Linien nicht zueinander parallel sind).

Das bedeutet: die Dauer als Merkmal für Vokalhöhe ist *teilweise geplant* = der Sprecher produziert absichtlich Dauerunterschiede grob zwischen hohem = /i, e/ und tiefem = /ε, a/, um dem Hörer zu vermitteln: hier ist ein hoher/tiefer Vokal (man verwendet aber nicht absichtlich die Dauer um innerhalb dieser Kategorien zu differenzieren: hier sind die Dauerunterschiede zwischen /i, e/ oder zwischen /ε, a/ eher mechanistisch). Daher S. 629: 'In sum, the results suggest that a slower rate increases the duration of low and low-mid vowels (= /ε, a/) by a greater amount than that of high and high-mid vowels (= /i, e/).

Den Rest dann weglassen/überfliegen bis 5.2.3. American English (S. 630):

'Indeed, the results in Table 11(a) show a significant vowel height x rate interaction for all speakers.' (S. 634)

/i:/ und /e/ (= Vokal in 'say') sind gespannte Vokale.

Hier kam es zu einer Interaktion zwischen Vokalhöhe und Sprechgeschwindigkeit. Das bedeutet, dass der Sprecher /i:/ absichtlich kürzer macht, um es von /e/ zu differenzieren.

Das gleiche gilt für /ɪ, ε, æ/ ('bit', 'bet', 'bat'). Hier gibt es auch eine Interaktion (das sieht man nur teilweise in Fig. 3, weil die Linien zueinander nicht ganz parallel sind). Daher planen Sprecher, /ɪ/ kürzer zu produzieren als /ε/, und /ε/ kürzer zu produzieren als /æ/.

Daher S. 636: 'The interaction between vowel height and rate is compatible with the hypothesis that American English speakers manipulate the durational differences associated with vowel height as a secondary cue to vowel identity.'

[manipulate = planen]

Den Rest überfliegen/weglassen/vergessen bis S. 641: Intrinsic vowel duration: perception experiment:

Hier wird weiterhin geprüft, ob die gefundene Planung in der katalanischen Vokalhöhe tatsächlich wahrnehmbar ist.

d.h. im Produktionsexperiment (siehe oben) wurde festgestellt, dass Katalanisch-Sprecher planen, tiefe Vokale länger als hohe Vokale zu produzieren. Wenn dies so ist - und wenn das zum Zweck der Hörerverständlichkeit verwendet wird (also um den Unterschied tief/hoch besser zu erkennen), dann müsste das auch wahrnehmbar sein. Dies wird hier getestet.

Die Dauer von einem Vokal, der ambig zwischen /e/ und /ε/ war, wurde manipuliert (nach oben und nach unten gesetzt) in einer /bV/ Silbe. Die Stimuli wurden Hörern

einzel präsentiert. Sie mussten dann in einer vierfachen Wahl entscheiden, was sie hörten: entweder: ve /be/ 'he comes', oder bè /bɛ/ 'sheep' oder vi /bi/ 'wine', oder va /ba/ 'he goes') - S. 643

Das wichtigste hier ist die Abbildung 4 auf S. 644. Dies zeigt die Identifizierungsrate (in %) für /e/ als Funktion der Dauer.

Wenn Hörer die Dauer als akustisches Merkmal für den /e, ɛ/ Unterschied nicht verwenden, dann würden wir eine flache Linie erwarten - und dies ist nicht der Fall.

Auf der anderen Seite erwähnt Solé nicht: die Kurve ist nicht gerade als eine Sigmoidalfunktion (S-Kurve) zu beschreiben (siehe Statistik - man bekommt einen Umkipppunkt und eine Sigmoidalfunktion, wenn es sich um den Unterschied zwischen zwei phonologische Kategorien handelt).

Daher die Schlussfolgerung (S. 645): 'Catalan speakers can hear durational differences in vowels and they use these differences in vowel categorisation. Thus a physiologically-based difference (such as intrinsic vowel durational differences) may become part of a perceptual cue distinguishing different vowel heights.'

Den Rest kann man weglassen bis zur Diskussion (S. 646):

Das Wichtigste aus der Diskussion:

'Returning to our original question of whether or not secondary features to contrasts are under the control of the speaker or merely result from implementational factors, the results suggest that differences in vowel duration are actively manipulated by the speaker and exploited as a cue to vowel height differences in some languages (like American English and Catalan), whereas they are by-products of the implementation of the height contrast (i.e., differences in jaw displacement) in other languages (like Japanese).

## # 8 Dissimilation

#####

Ich würde mit einigen Daten aus ohala84c, ohala89 als Einleitung beginnen. D.h., in der Einleitung eventuell verschiedene Beispiele aus der Dissimilation darstellen, dann Ohalas Theorie.

Dazu gerne die Folien nehmen/ändern aus L1.pptx, S. 17-21 in Verbindung mit ohala84c (S. 8-10).

Dann alderete06.

Dann die 2 Experimente in Abrego-Collier und Harrington.

ohala84c

+++++

/b̥end/ wird zu /bend/ ([..] bedeutet 'behauchte Stimme'). Der Hörer meint, die behauchte Stimme im /b̥/ kommt aufgrund der Koartikulation wegen des behauchten /d/ vor, und rechnet es perceptiv heraus - der Lautwandel entsteht also wegen eines Hörerfehlers.

Der Ursprung der Dissimilation kann nicht verloren gehen. D.h. das zweite /w/ bleibt im Lateinischen /kinkwe/ nach der Dissimilation bestehen.

Dissimilation kommt hauptsächlich bei Lauten vor, die ein sehr langes Zeitfenster haben - d.h. die einen Einfluss auf mehrere - meistens davor kommende - Konsonanten oder Vokale haben. Dies betrifft u.a. Lippenrundung, aber insgesamt oft die sekundäre Artikulation (Palatalisierung, Velarisierung, usw).

Dissimilation ist eher antizipatorisch, genauso wie Koartikulation auch. Dies ist ein weiterer Grund, weshalb die Koartikulation die Grundlage einer Theorie der Dissimilation sein soll.

ohala89.pdf, §3.2.3

+++++

Hier werden einige klassische Beispiele der Dissimilation erläutert, und dann auch erklärt, wieso die 'conditioning environment' in einer Dissimilation nicht verloren gehen kann. Die berühmtesten Beispiele sind Latein /kwinkwe/ -> /kinkwe/ und griechisch /thrikhos/ -> /trikhos/.

alderete06.pdf S. 1-9

+++++

'Dissimilation has been shown to be statistically systematic when observed over lexical items' ist nicht uninteressant. Dies bedeutet, dass einige wiederholte Konsonanten wie /spap/ usw. im Deutschen relativ selten sind. Siehe dann auch das Beispiel auf S. 2/3 zum Arabischen.

Das Beispiel aus der Sprache Berber (S. 2): Ein Präfix /m/ wird zu /n/, wenn im Stamm ein labialer Konsonant vorkommt. Das ist eigentlich ganz interessant, denn man sieht,

dass das Präfix /m/ zu /n/ dissimiliert, und zwar sogar in xalaf (m-xalaf -> n-xalaf), d.h. in einem Stamm /xalaf/, in dem der labiale Konsonant /f/ final ist.

Die Tabelle 2 kann überflogen werden - oder auch kurz erwähnt. Diese zeigt, dass Dissimilation einen Einfluss auf das Lexikon hat; z.B. in einigen Sprachen 'mögen' Wörter nicht so gerne, wenn aneinandergrenzende Konsonanten die selben Merkmale haben - wie labial gefolgt von labial (daher der Eintrag von 0 in der Tabelle bei Labial x lab: das bedeutet, dass die Reihenfolge von labial +labial im arabischen Lexikon gar nicht vorkommt).

Die Tabelle auf S. 4 zeigt, welche Lautklassen am häufigsten von der Dissimilation betroffen sind. Eventuell nur kurz erwähnen, dass Liquide, laryngeale Laute, und Laute mit derselben Artikulationsstelle (Place) am meisten von der Dissimilation betroffen werden.

Danach - bis 16.3 - reicht es, Folgendes sehr knapp zusammenzufassen:

In manchen Sprachen kommt die Assimilation nur unter gewissen Voraussetzung vor, und zwar wie folgt: Es wird hier von *target* und *trigger* gesprochen.

z.B. Latein kwinkwe -> kinkwe. Das erste /w/ ist das Target (= wird von der Dissimilation betroffen), das zweite /w/ ist der Trigger (= der Auslöser der Dissimilation).

(a) Root adjacency = in manchen Sprachen kommt die Dissimilation nur vor, wenn Target/Trigger unmittelbar aufeinander folgen, z.B. Arabisch hat ein gerundetes /g/ in /gwrđ/, das nach einem /m/ in am-grđ entrundet wird, usw.

(b) Syllable adjacency: 'the target and trigger must be contained in adjacent syllables' wie gesagt, das ist wenn target und trigger in aufeinanderfolgenden Silben sind

(c) Domains: 'the target and trigger must fall within a specified domain, e.g., a morphological domain (root, stem, word, morpheme) or prosodic domain (e.g., syllable, foot, rime)'.  
Hier ist die Domäne noch größer, d.h. es reicht wenn target und trigger sich im selben Wort oder Morphem befinden. Siehe das obige Beispiel in Berber m-xalaf, in dem das Target /f/ und der Trigger /m/ im selben Wort erscheinen.

(d) No intervener : 'for target segment x and trigger y, there must be no segment z, of a specified type, that intervenes between x and y'

etwas langwierig ausgedrückt vielleicht an dem Beispiel zeigen d.h.

lun+alis -> lunaris aber flor+alis -> floralis und nicht froralis (hier ist x das erste /l/, z = das /r/, y = das /l/ in -alis)

S. 6/7: eine Zusammenfassung von Ohalas Theorie und auf S. 8 werden viele Beispiele erläutert, die mit Ohalas Theorie konsistent sind.

Das Beispiel (5) S. 8:

stoja, [a] (also nah an Kardinalvokal 4) ist ein vorderer Vokal. Wenn Dissimilation vorkommt, meint der Hörer, dass das davorkommende /j/ der Ursprung für das vordere [a] ist (also dass die Vorverlagerung nur wegen des /j/ zustande kommt). Dann rechnet

inkorrekt der Hörer diese Vorverlagerung aus dem vorderen [a] raus, und bekommt ein rückverlagertes [a], also Kardinal Vokal 4 -> Kardinal Vokal 5.

Sonst eventuell nur noch 2 Sachen aus diesem Paper erwähnen:

Auf S. 8 wird das folgende Problem für Ohalas Modell dargestellt: wieso werden einige Laute von Assimilation und andere von Dissimilation betroffen? Z.B. ist Assimilation häufig in z.B. /in/ -> /im/ (impossible), aber es gibt kaum Fälle wo /imp/ zu /inp/ dissimiliert. Umgekehrt ist Dissimilation von Liquiden häufig, aber es gibt kaum Liquid-Assimilation (/r/-> /l/ vor /l/ z.B.). Es ist nicht eindeutig, ob Ohalas Theorie diese Asymmetrie erklären kann.

S. 9 Hier wird versucht, die Fehler in Zungenbrechern mit dem Auslöser für Dissimilation zu verbinden, d.h. hier wird die Frage gestellt, ob Dissimilation vorkommt, weil aufeinanderfolgende, ähnliche oder identische Lautreihenfolgen oft zu Sprechfehlern führen wie in 'Peggy Babcock' oder 'unique New York'.

Der Rest vom Artikel kann weggelassen werden.

Abrego-Collier

+++++

Ein 7 stufiges Kontinuum wurde synthetisiert zwischen /l/ und /r/. X ist der betroffene Laut, Y ist der Trigger im Sinne von oben.

Das Kontinuum wurde eingebettet in der Position X in verschiedenen Logatomen der Form aXaYa und aYaXa für Y = /d, l, r, dn, ln, rn/. Wenn Y ein /l/ enthält, dann müsste es in X laut Ohalas Theorie mehr /r/-Antworten geben; wenn Y ein /r/ enthält, dann mehr /l/-Urteile. Dieser Effekt müsste antizipatorisch größer sein in aXaYa als progressiv in aYaXa (progressiv, weil wegen der Koartikulation Y X beeinflussen müsste).

Die Ergebnisse zeigten keine Evidenzen dafür - eher sogar im Gegenteil: In aXaYa und aYaXa gab es mehr /l/ Antworten, wenn Y ein /l/ enthielt. Dies ist deutlich in Abb. 1. Laut Ohalas Theorie müssten rot und blau umgekehrt sein.

Daher: 'The prediction that the presence of /r/ should effect a higher rate of /l/-identification for the target liquid in Hypothesis A was not borne out, as conditioning /r/ did not have a significant effect for this study. Meanwhile, Hypothesis B predicted that the occurrence of /l/ should cause listeners' /r/-/l/ category boundary to shift toward more responses of /r/ in identifying the continuum liquid. A significant effect of conditioning /l/ was found, but the empirical data supports the opposite of what was expected.'

Viel mehr als das braucht man von diesem Paper nicht erwähnen. Aber was nicht ganz klar ist, ist, wieso ein /l/ als Trigger sogar zu einer größeren Wahrnehmung von /l/ in X führen müsste.

harrington16.pdf

+++++

Das Experiment befasst sich mit der Wahrnehmung vom italienischen 'canto' und 'quanto', /kanto, kwanto/

Laut Ohala müsste man eher 'canto' perzipieren in:

*ho detto canto/quanto quattro volte*

im Vergleich zu

*ho detto canto/quanto sette volte*

Das sollte so sein, weil laut Ohala das [w] von 'quattro' das [w] in 'quanto' perzeptiv maskiert, d.h. Hörer werden das [w] in 'quanto' teilweise als Folge der antizipatorischen Koartikulation aufgrund des [w] in 'quattro' interpretieren, und es perzeptiv herausrechnen. Das wird aber nicht in 'ho detto canto/quanto sette volte' passieren, da kein [w] in 'sette' vorkommt.

Die zweite Hypothese betrifft Akzentuierung.

Solange der Hörer imstande ist, den Trigger (Lippenrundung im zweiten [w]) vom Target (das ersten [w]) zu differenzieren, dann wird es keine Dissimilation geben.

Dissimilation kommt eher zustande, wenn Target/Trigger perzeptiv voneinander nicht differenziert werden können (weil dann weiß der Hörer nicht mehr: kommt diese Lippenrundung wegen des Targets (dem ersten [w]) oder des koartikulatorischen Triggers (dem zweiten [w]) vor?

Unsere Hypothese ist, dass es NOCH schwieriger wird, Target und Trigger in deakzentuierter/hypoartikulierter Sprache voneinander zu differenzieren in:

*ho DETTO canto/quanto quattro volte*

(DETTO = nuklearakzentuiert, canto/quanto deakzentuiert)

als in

*ho detto CANTO/QUANTO quattro volte*

Der Test

Ein Kontinuum zwischen canto/quanto wurde synthetisiert und in die 4 Trägersätze (gelisted auf pdf S 6 (s. 20)) eingebettet. Hörer mussten in einem two-alternative forced-choice -Test immer beurteilen, ob sie entweder 'canto' oder 'quanto' hörten.

Die Gruppenergebnisse sind in Abbildung 1 auf pdf S. 8.

Auf der x-Achse: stimulus 1 = eindeutiger 'canto', stimulus 11 = 'eindeutiger quanto'  
grau = '\_ sette', schwarz = '\_ quattro'

1. Laut der Hypothese in Ohala müsste man mehr 'quanto' wahrnehmen im Kontext von \_sette

Aber genau das Gegenteil kam raus! D.h. man hört mehr 'quanto' im Kontext von '\_quattro'. Dieses Ergebnis ist übrigens sehr ähnlich wie das von Abrego-Collier oben - er fand mehr, nicht weniger /l/-Wahrnehmungen vor einem danach kommenden /l/.

2. Man müsste mehr 'canto' wahrnehmen im deakzentuierten Kontext d.h. mehr 'canto' in

*ho DETTO canto/quanto quattro volte*

als in

*ho detto CANTO/QUANTO quattro volte*

Das ist so - aber das war AUCH so im Kontext von '\_sette'. d.h. unabhängig davon, was der Kontext ist (ob 'quattro' oder 'sette') nimmt man mehr 'canto' im deakzentuierten Kontext wahr.

Unsere Erklärung ist wie folgt. In beiden Kontexten war F2 in '-anto' von 'canto/quanto' sehr niedrig, d.h. der Hörer hat Lippenrundung in 'canto' und 'quanto' wahrgenommen, egal ob der danachkommende Wort 'sette' oder 'quattro' war.

Im deakzentuierten Kontext kann der Hörer so gut wie gar nicht zwischen der Lippenrundung im '-anto' und dem [w] in 'quanto' differenzieren; und hört daher im allgemeinen nur 'canto' (oder sehr viel 'canto') im deakzentuierten Kontext.

Daher sind die gestrichelten psychometrischen Kurven (deakzentuiert) beide links verlagert im Vgl. zum akzentuierten Kontext.

Siehe dann auch pdf S.15 für eine Zusammenfassung der Hypothese bez. Akzentuierung und auch Erklärung der Ergebnisse.

Mehr von dem Artikel muss nicht erwähnt werden.

## #9 Sound change and the covariation of speech

beddor07.pdf

+++++

Einleitung. Hier handelt sich um die Beziehung zwischen der Produktion und Perzeption von Koartikulation. Hörer perzipieren die koartikulatorischen Verhältnisse im Signal auf eine ähnliche Weise, wie sie in der Produktion der Sprache vorkommen. Da z.B. Nasalisierung im Vokal in Sprachen wie Englisch und Deutsch wegen Kontext entsteht (in /ÑN/ Kontexten), rechnen Hörer die kontext-bedingte Nasalisierung [ ~ ] aus dem Signal heraus (perzeptuelle Kompensation von Koartikulation). Ohalass Idee ist, dass Hörer das manchmal nicht tun, wodurch die Nasalisierung dem Vokal zugeordnet wird, sodass die Nasalisierung im Vokal phonologisiert wird (Nasalisierung wird als Bestandteil des Vokals aufgefasst, um z.B. den Unterschied 'seen'/'seed' wahrzunehmen).

Neuere Untersuchungen von Beddor und anderen zeigen Folgendes:

1. Es gibt sehr viel Variabilität zwischen Hörern, inwiefern sie für die kontext-bedingte Nasalisierung kompensieren (= Nasalisierung aus dem Signal heraus rechnen).
2. Das Ergebnis von Diskriminationsexperimenten: Wenn Hörer die koartikulatorische Nasalisierung aus dem Signal heraus rechnen, dann dürfte der Hörer keinen Unterschied zwischen  $\tilde{V}$  und V wahrnehmen, wenn  $\tilde{V}$  in N\_N vorkommt. Das ist aber nicht ganz der Fall - und das zeigt, dass zumindest ein Teil der kontext-bedingten Nasalisierung am Vokal hängenbleibt.

Was also bei Beddor neu dazu kommt (im Vgl. zu Ohala):

'Having contextual residue on the target sound means that change might take place even when the coarticulatory source is detected.'

D.h., auch wenn Hörer die Quelle der Koartikulation (ein finales /m/ oder /n/) richtig erkennen, wird trotzdem ein Teil der Nasalisierung mit dem Vokal perzeptiv assoziiert. Dies ist der Auslöser für Lautwandel und erklärt auch, weshalb sich in einigen Sprachen die kontrastive Vokalnasalisierung wie im Französischen entwickelt hat, aber ohne dass die Quelle wie im Französischen komplett getilgt wurde.

9.2.

Contextual covariation (9.2.1) in der Sprachproduktion:

Dies wird am besten durch Abb. 9.1 gezeigt; Kovariation bedeutet, dass der Anteil der Nasalisierung im Vokal im umgekehrten Verhältnis zur Dauer des Nasalkonsonanten steht. Je mehr Vokalnasalisierung auftritt, desto kürzer wird der danach kommende /n/ sein, usw. . Dies ist ganz deutlich in Amerikanisch Englisch für finale stimmlos/ stimmhaft Paare zu erkennen: der Vokal in 'sent' ist fast komplett nasalisiert und das /n/ fast komplett getilgt; stattdessen ist für 'send' nur der letzte Teil vom Vokal nasalisiert, und es gibt ein viel längeres /n/ als in 'sent'.

Sprachperzeption.

Wie wird diese Kovariation in der Produktion behandelt? D.h., welche Strategien werden entwickelt, um VN (wenig Vokalnasalisierung, langer finaler Nasalkonsonant) mit  $\tilde{V}n$  (viel Vokalnasalisierung mit kurzem Verschluss für den Nasalen) perzeptiv gleich zu setzen? D.h. es soll für den Hörer egal sein, ob Nasalisierung im wesentlichen im

alveolaren Verschluss teil oder im Vokal vorkommt - die Hauptsache ist, dass Nasalisierung irgendwo im Reim wahrgenommen wird.

Das Experiment: Es wird untersucht, ob ein 'perceptual trading relationship' zwischen den Cues vorliegt. D.h., wenn zwei Cues A und B beide eine phonologische Klasse signalisieren, dann ist, je mehr sich ein Hörer auf A verlässt, B umso weniger wichtig - und umgekehrt, wenn ein *perceptual trading relationship* vorliegt.

Akustische Stimuli:

Der Vokal, V, bestand aus einem oralen gefolgt von einem nasalisierten Teil. Der nasalisierte Teil war entweder kurz:  $\tilde{V}S$  oder lang  $\tilde{V}L$

Die Dauer vom danach kommenden /n/ wurde in 8 Schritten manipuliert.

Es gab dann 3 Sorten von Paaren A ~ B

*N-only* =  $\tilde{V}SNS \sim \tilde{V}SNL$

d.h. der nasalisierte Teil war immer kurz und das /n/ entweder kurz (A) oder lang (B)

*Cooperating* =  $\tilde{V}SNS \sim \tilde{V}LNL$

d.h. in A waren der nasalisierte Teil sowie /n/ kurz und in B waren beide lang

*Conflicting* =  $\tilde{V}LNS \sim \tilde{V}SNL$

d.h. in A war der nasalisierte Teil lang und /n/ kurz; in B anders herum

Hypothese 1:

Die Diskriminierung in *Conflicting* müsste schwierig sein, *wenn es ein Trading Relationship gibt*.

Daher

1.1. *Conflicting* < *Cooperating*

1.2. *Conflicting* < *N-only*

"<" bedeutet: schlechtere Diskriminierung

(wenn ein Trading Relationship vorhanden ist)

Hypothese 2:

Hörer schenken dem Auftreten der Vokalnasalisierung in einem -/t/ Kontext viel Aufmerksamkeit. Das ist so, weil Vokalnasalisierung der wesentliche Cue in Wörtern wie 'bent' ist.

Daher müsste der Unterschied zwischen Paaren, die sich in dem Anteil der Vokalnasalisierung unterscheiden, in einem /t/ viel salienter als in einem /d/ Kontext sein.

Eventuell kommt das Trading Relationship auch eher in /d/ zustande. Vor /t/ ist die Nasalisierung im Vokal eher entscheidend.

Daher

2.1 *Cooperating* /d/ < *Cooperating* /t/

(Die Diskriminierung der *Cooperating*-Paare in einem /d/ Kontext ist schwächer als in einem /t/ Kontext)

2.2 *Conflicting* /d/ < *Conflicting* /t/

2.3. *Conflicting* /t/ ähnlich wie *Cooperating* /t/, weil sich in beiden Fällen die Paare im Anteil der Vokalnasalisierung unterscheiden.

Die Ergebnisse

1.1. ja im /d/ Context

1.2. nein

2.1 ja

2.2 ja

2.3 ja

Bezüglich 1.2. (*Conflicting* < *N-only*): Man erwartet eine sehr schlechte Diskriminierung in  $\tilde{V}LNS \sim \tilde{V}SNL$  (*conflicting*) wenn im /d/-Kontext ein Trading Relationship vorhanden ist. Aber dies war nicht für alle Hörer der Fall (Abb 9.5 zeigt *Conflicting* < *N-only* für 10 Hörer, aber nicht für 12 Hörer).

Die wichtigste Schlussfolgerung

1. In einem Kontext wie 'bend' hört man kaum den Unterschied zwischen langer Vokalnasalisierung + kurzem /n/ vs. kurzer Vokalnasalisierung + langem /n/. Es gibt also ein Trading Relationship. Dies entspricht in der Produktion der Sprache einer variablen Alignierung der Velumöffnungsgeste ca. wie folgt:

Velum-Senkung = ++++ und überlappt viel mit dem oralen Teil des Vokals --- , aber dann ist der alveolare Verschluss /n/ kaum nasalisiert und der orale Verschluss /d/ ist lang:

```
+++++
-----nnnnnnnnndddddddddd
```

Velum-Senkung überlappt kaum mit dem oralen Teil des Vokals und dann ist /n/ lang und der orale Verschluss für /d/ kurz:

```
+++++
-----nnnnnnnnnnnnnnndddd
```

2. Die Stärke des perceptual trading relationships variiert zwischen Hörern.

3. Die Stärke des *perceptual trading relationships* ist viel größer vor /d/ als vor /t/. Im stimmlosen Fall hört man die Dauer der Vokalnasalisierung (als Cue für /n/) viel deutlicher.

Die Abbildungen in 9.7 sind jetzt nicht wichtig.

Viel wichtiger ist die Interpretation der Daten:

Für Beddor ist, was in 'bend' passiert (trading relationship), die erste Phase des Lautwandels, in dem schließlich VN zu  $\tilde{V}$  wird (Latein manus -> Fr. main = m $\tilde{e}$ )

Die zweite Phase ist, was man in 'bent' hat: also lange Nasalisierung, auf die Hörer deutlich reagieren, aber wo doch noch ein kleines /n/ übrig bleibt. D.h. 'a predictable property' (= Koartikulation) 'has become distinctive' (= perzeptiv salient, oder ist wie man in 'bent' sieht, auf dem Weg salient zu werden, viel salienter als in 'bend').

beddor12.pdf

+++++

Koartikulation ist meistens kohärent (Kohärenz = Zusammenhang), weil überlappende Gesten für den Hörer wichtige Informationen zum phonetischen Inhalt übertragen. Z.B. wenn der Kiefer im /ɪ/ von 'Mittag' wegen dem /a/ in der zweiten Silbe gesenkt wird (wodurch F1 höher wird), dann hat der Hörer Informationen zu dem /a/ schon in der ersten Silbe (kann also schneller das /a/ entschlüsseln).

Koartikulation kann perzeptive Urteile manchmal ambig machen - wie wenn der Hörer nicht genau weiß, ob ein /t/ in 'eins' vorkommt oder nicht.

Es ist für Beddor die erste Variante und dessen Beziehung zu Lautwandel, die hier wichtig ist (für Ohala ist es eher die zweite Variante).

Abb. 1 zeigt wieder einige Ergebnisse aus Beddor (2007):

- dass die Dauer der Vokalnalisierung und Nasalverschluss ko-variiieren (je länger das eine, desto kürzer ist das andere, und umgekehrt).
- dass dieser Effekt noch größer ist vor /t/ als vor /d/.

S. 42 und 43: Dies ist eine Erweiterung von Beddor (2007) dadurch, dass es sich um forced-choice Unterschiede handelt, d.h. Hörer mussten für verschiedene Stimuli identifizieren: 'bet' oder 'bent'; 'bed' oder 'bend'.

Die Hypothesen:

1. Die Wahrscheinlichkeit, dass man einen Nasalkonsonanten mit zunehmender Nasalisierung hört, ist größer in 'bet-bent' als in 'bed-bend' (weil Hörer schneller auf Vokalnalisierung aus all den in beddor07 erwähnten Gründen reagieren).
  2. Es gibt ein Trading Relationship - je kürzer die Dauer vom /n/ umso länger muss die Dauer der Vokalnalisierung sein, um ein /n/ wahrzunehmen.
- 1 und 2 sind in Abb. 2 bestätigt.

Man sieht auch auf der nächsten Seite in Abbildung 3, dass es hier viel Variation zwischen den Hörern gab. Hier ist diese Zeile wichtig: 'thus different listeners systematically accessed a given lexical item through different acoustic information'.

Es ist deutlich (S.45), dass Hörer sich also darin unterscheiden, wie die Überlappung perzeptiv verarbeitet wird, um ein Mapping zwischen Signal und Lexikon zu ermöglichen, aber warum das so ist, ist nicht ganz klar. Eine Möglichkeit laut Exemplartheorie ist, dass Hörer sich wegen Erfahrungsunterschieden mit der Sprache unterscheiden. D.h., laut Exemplartheorie werden Sprachsignale im Gedächtnis gespeichert. Ein Wort und die dazu passende Phonologie ist eine statistische Verteilung über diese Sprachsignale im Gedächtnis. Da Personen sich mit unterschiedlichen Sprechern unterhalten, variieren notwendigerweise die gespeicherten Signale und daher die Beziehungen zwischen Lexikon, Phonologie, und erinnerten Sprachsignalen. Eventuell kommen diese Unterschiede in Abb. 3 usw. deswegen zustande.

Jedoch entstehen diese Unterschiede nicht alle, weil die Beziehungen zwischen Wörtern und Sprachsignalen sprecher- (hörer-) spezifisch ist:

1. Weil man das gleiche Ergebnis in der Hörervariation mit Logatomen bekommt (Abb. 4)
2. Diese Hörerunterschiede in Abb. 4, die sich mit /mb/ befassen, nicht wegen Erfahrungen entstanden sein können, weil /m/ immer deutlich in Wörtern wie number, jumbo usw. produziert wird.

Real-time processing of coarticulated speech (S. 49):

Hier handelt sich um ein eye-tracking Experiment, das feststellen soll, ob sich Hörer in der zeitlichen Aufnahme von Koartikulation unterscheiden.

Eye-tracking: Die Augenbewegungen werden gemessen. Hörer sehen einen Bildschirm mit Bildern von 2 Alternativen z.B. 'set' vs 'sent'. Siehe auch Abb. 1 in beddor15.pdf.

Sie hören dann einen akustischen Stimulus (z.B. synthetisch manipulierte Stimuli der Wörter). Man misst, wie schnell sich die Augenbewegungen in Richtung des einen oder des anderen Worts ausrichten.

Abb.5: Wenn der Vokal nasalisiert ist, und man das /n/ wegschneidet, dann identifiziert man 'sent' schneller als 'send' (Abb. 5 rechts); dies ist konsistent mit der Hypothese, dass die Vokalnalisierung ein deutlicherer Cue für /n/ ist in 'sent' als in 'send'. Abb 6. zeigt, dass es sehr viel Variation zwischen den Hörern gibt, inwiefern die Vokalnalisierung als Cue für einen Nasalkonsonanten getrackt wird.

Schlussfolgerung

-----

Wenn es ein Trading Relationship zwischen Cues gibt - wie zwischen Vokalnalisierungsdauer und Dauer des /n/-Verschlusses - dann kann es passieren, dass Hörer die Wichtigkeit dieser Cues unterschiedlich einstufen. Einige verlassen sich mehr auf das Vorhandensein von /n/, andere darauf, dass Vokalnalisierung vorkommt. Das ist der Hauptgrund, weshalb die Ergebnisse Hörervariationen zeigen.

Damit das zu Lautwandel wird, müssten diese Unterschiede in der perzeptiven Verarbeitung auf die Sprachproduktion übertragen werden. Aber solche Studien gibt es noch nicht. D.h. wir wissen nicht, ob Hörer, für die die Nasalisierung im Vokal wichtiger ist als der /n/-Verschluss, solche perzeptive Effekte auch in der Produktion duplizieren (d.h. ob ein solcher Hörer 'sent' mit viel Nasalisierung und kaum einem /n/ produziert usw).

Dies ist aber die Forschungsfrage in beddor15.pdf:

beddor15.pdf

+++++

Die Hauptfrage hier ist:

'Do a language user's articulatory strategies correlate with that user's particular perceptual weightings?'

und

'do language users who produce more innovative variants' (z.B. hauptsächlich Vokalnalisierung in 'send') 'also weight the innovative property more heavily in perception?'

Der erste Teil ist eine wiederholte Zusammenfassung des Perzeptionsexperiments mit eye-tracking.

Für dieselben Hörer wurde der zeitliche Anteil der Nasalisierung im Vokal ermittelt. Der akustische Parameter, um dies zu tun, ist das Verhältnis  $A1/P0$  (entspricht  $A1 - P0$  auf der logarithmischen dB-Skala, da  $\log(A/B) = \log(A1) - \log(P1)$ ), wobei  $A1$  der Amplitude des ersten Formanten und  $P0$  der Amplitude des tiefsten Nasalformanten entspricht.

Je nasalierter, desto kleiner  $A1$  und desto größer  $P0$ , und daher desto kleiner  $A1/P0$  und daher desto negativer  $A1 - P0$ . Daher liegt für CVD die Kurve oberhalb der Kurve von CVNT in Abb. 2 links.

Abb. 2 zeigt Folgendes. Vielleicht eher nur auf Reihen 2 und 3 konzentrieren.

Reihe 2: Der Hörer schaut ganz deutlich auf 'bent' schon ab Beginn des nasalisierten Vokals. d.h. sobald E07 etwas nasalisiertes hört, wird es interpretiert als 'bent' (und nicht 'bet'). Kompatibel damit: E07 hat ganz deutlich nasalisierte Vokale in Wörtern wie 'bent' (Abb 2, Reihe 2 links). E07 ist daher ein innovativer Hörer, weil - ähnlich wie im Französischen - für diese Versuchsperson der Nasal durch Nasalisierung im Vokal übertragen wird.

E02 ist dagegen viel konservativer. E02 differenziert in der Perzeption nur zwischen 'bent'/'bet' ab dem Beginn von [n]. Die Sprachproduktion von E02 ist damit kompatibel, da es in dessen Sprachproduktion kaum Nasalisierung im Vokal gibt.

Afrikaans: Es handelt sich hier um einen Lautwandel-im-Fortschritt aus dem Bereich der Tonogenese, also wo [pa, ba] -> [pa\, pa/] (fallend vs. steigend) wird.

Es soll festgestellt werden, ob Hörer, die an dem Lautwandel in der Sprachproduktion teilnehmen, für den phonologischen Kontrast in der Perzeption der Grundfrequenz mehr Aufmerksamkeit schenken.

Abb. 1 zeigt Produktionsdaten für jung (die eher am Lautwandel teilnehmen) vs. alt (die kaum daran teilnehmen).

Für alt ist VOT viel ausgeprägter als für jung; aber die Gruppen unterscheiden sich nicht in der Produktion von  $f_0$ : in beiden Gruppen wird /p/ durch einen höheren  $f_0$ -Wert im Vgl. zu /b/ differenziert.

Dann ist die Frage, ob jung in der Perzeption mehr Gebrauch von  $f_0$  macht, um /ba, pa/ voneinander zu differenzieren.

Es gab 3 Sorten von Kontinua für /bas-pas/ und /dur-tur/

In allen drei wurde  $f_0$  progressiv angehoben

(a) zusätzlich gab es deutliche VOT Unterschiede

(b) diese Unterschiede wurden teilweise reduziert

(c) es gab keine VOT Unterschiede

Die Ergebnisse vom forced-choice Perzeptionstest (die Hörer mussten pro Stimulus antworten: bas oder pas (dur oder tur) zeigen, dass ältere Hörer in (a, b) VOT mehr Aufmerksamkeit als Cue schenken. Jung tut das nicht, weil  $f_0$  für jung ausreicht, um in (a, b) zwischen /bas-pas/ oder /dur-tur/ zu differenzieren, wenn  $f_0$ -cues vorhanden sind.

Insofern gibt es einen Match/ eine Übereinstimmung zwischen den Modalitäten (zwischen Produktion und Perzeption der Sprache). Alt macht von VOT in der Produktion der Sprache mehr Gebrauch als jung, und schenkt VOT mehr Gewicht in der Perzeption als jung. Solche Beziehungen sind auch für Individuen erkennbar (Abb 5). Z.B. A01 verwendet VOT um /b,p/ in der Produktion zu differenzieren, und reagiert sehr stark auf VOT in der Perzeption; andersherum für A23: hier ist VOT in dessen Produktion irrelevant für den /b, p/ Unterschied, und A23 reagiert viel weniger auf VOT als Cue für den /b, p/ Unterschied in der Perzeption.

#### Diskussion

Es wurde erwartet, dass  $f_0$  als Cue variabel ist zwischen Hörern, ähnlich wie Hörer die Nasalisierung auf eine variable Weise für 'bent' vs 'bet' einsetzen. Aber das ist nicht der Fall.  $f_0$  ist schon ein sehr prominenter Cue für den /b, p/ Unterschied. Die Variabilität ist eher in VOT. Dies deutet darauf hin, dass der Lautwandel schon ziemlich fortgeschritten ist.  $f_0$  ist bereits komplett in der Perzeption phonologisiert worden, und VOT entfällt, oder wird unwichtiger als Cue (dadurch dass junge Hörer im Vergleich zu älteren viel weniger Gebrauch von VOT machen).

Abb. 7. Meistens sind Produktion und Perzeption im Einklang miteinander. Es wird kaum der Fall sein, dass man einen Cue in der Produktion verwendet, und dann darauf in der Perzeption nicht reagiert (unten rechts). Es kann laut Beddor wohl sein, dass man einem Cue in der Perzeption (wie  $f_0$  in ba, pa oder Nasalisierung im Vokal) Aufmerksamkeit schenkt, ohne dass dieser Cue auf eine reliable Weise in der Produktion umgesetzt wird. Dies deutet also auf eine Interpretation hin, dass Lautwandel zuerst aufgrund von Innovationen in der Perzeption entstehen; Änderungen in der Produktion kommen erst danach.

## #10. Sound change and the evolution of speech

### 1. Was ist Lindbloms H&H Theorie?

-----

Aus [1]. Speech is 'adaptively organised' d.h. sehr viel Variabilität in der gesprochenen Sprache kommt zustande, weil sich der Sprecher an die Sprechsituation anpasst. Der Sprecher produziert eine klare, artikulatorisch aufwendige Sprache (=hyperarticulation), wenn vom Sprecher berechnet wird, der Hörer wäre kaum imstande, die gesprochene Sprache aus dem Kontext vorherzusagen. Z.B. wenn man jemanden vorstellt, kann der Hörer von dem Namen keine Ahnung haben: hier setzt der Sprecher die hyperartikulierte Sprache ein. Wenn jedoch die Äußerung oder Teile davon aus dem Kontext berechenbar sind, dann kommt es zu Verschleifungen/Reduzierungen usw., also zu einer hypoartikulierte Sprache.

Hyper- und Hypoartikulation ist also ein 'adaptation to the environment' (Anpassung an die Umgebung). Lautwandel könnte 'adaptive' im evolutionären/Darwin'schen Sinne sein, weil Wandel aus einer sprachlichen Kommunikation entsteht, die sich optimal an die Umgebung/Bedürfnisse des Hörers anpasst. Aus dem Grund sind die häufigsten Formen von Lautwandel mit hyper- und hypoartikulierten Sprache verbunden. (Mehr Details zur H&H Theorie: lindblom88.pdf, S. 156-164)

Siehe auch lindblom95.rivling.pdf S.7 und8 für Beispiele der Hypo- und Hyperartikulation.

### 2. Die 'How' and 'why' Modi der Sprachperzeption

-----

Für Lindblom gibt es zwei Modi der Sprachperzeption: der Hörer konzentriert sich auf  
 (a) *was* (what) der Sprecher sagte  
 meistens im umgekehrten Verhältnis zu  
 (b) *wie* (how) er es sagte.

(a) kommt zustande, je mehr der Hörer auf den Kontext angewiesen ist, um die Bedeutung der Sprache zu entschlüsseln. z.B. Funktionswörter sind aus dem Kontext meistens vorhersagbar. Daher sind sie auch meistens hypoartikuliert (siehe 1 oben). Und genau hier setzt der Hörer seinen *was*-Perzeptionsmodus ein, und ignoriert zum großen Teil *wie* das Wort produziert wurde. Das ist auch gut so, da Funktionswörter aufgrund der Hypoartikulation sehr variabel/reduziert sind!

(b) wird eher eingesetzt, wenn der Hörer kaum Kontext verwenden kann - also in hyperartikulierter Sprache, die in der Sprachproduktion eingesetzt wird, wenn die Sprache nicht aus dem Kontext vorhergesagt werden kann. Hier ist aber glücklicherweise die Sprache sehr klar (hyperartikuliert), sodass Kontext kaum notwendig ist, um die Bedeutung zu entschlüsseln.

### 3. Bezug zu Lautwandel und Ohalas Modell

-----

Man kann sagen, dass sowohl in Ohalas als auch in Lindbloms Modell Lautwandel entstehen kann, wenn ausnahmsweise die Sprache vom Hörer *dekontextualisiert* wird.

Dekontextualisiert bedeutet: ein Laut wird ausnahmsweise nicht im Bezug zum Kontext dekodiert.

Dies kommt in Ohala vor, wenn ein Hörer für den Kontext nicht normalisiert (wie wir in den letzten Woche besprochen haben): z.B. wenn Vokal-Vorverlagerung nicht aus 'tut' herausgerechnet wird, dann perzipiert der Hörer /y/ statt /u/ - und das kann zu Lautwandel führen. Hier wird dekontextualisiert, weil der Hörer ausnahmsweise die Vokal-Frontierung nicht auf die Quelle der Koartikulation (das /t/) zurückführt.

Dekontextualisierung kommt in Lindblom vor, wenn ausnahmsweise ein *wie*-Modus falsch eingesetzt wird. Wie wir in 2. oben gesehen haben, wird der *was*-Perzeptionsmodus z.B. in Funktionswörtern eingesetzt, sodass der Hörer nicht das Signal, sondern den Kontext verwendet, um die Bedeutung wahrzunehmen. Sollte der Hörer aber einen *wie*-Perzeptionsmodus einsetzen, dann wird er die verschleifte, reduzierte phonetische Form der hypoartikulierten Sprache wahrnehmen, und die verschleifte/reduzierte Variante eventuell als mögliche Aussprache seinem Lexikon hinzufügen. In der Sprachproduktion könnte der Hörer dann auf diese neue Aussprache zugreifen - und das wäre dann Lautwandel.

Ein wichtiger Unterschied zwischen den Modellen ist wie folgt:

- In Ohala kommt Lautwandel vor, weil der Hörer ausnahmsweise zu wenig (z.B. Phonologisierung der Vokalnalisierung) oder zu wenig (Dissimilation) kompensiert.
  - In Lindblom kommt Lautwandel vor, weil ausnahmsweise dem Lexikon eine andere Aussprache hinzugefügt, und in der Sprachproduktion ausgewählt wird.
- Beide Modelle sind sich aber einig, dass Lautwandel aufgrund von Hörer-Dekontextualisierung vorkommt.

#### 4. Weitere Unterschiede zu Ohalas Modell: 1. Häufigkeit

-----

Die Teile von lindblom95.rivling.pdf 'extending the model' (Abs. 3.3 S. 18 und 3.4 S. 20) können weggelassen oder überflogen werden. Dann mit 3.5 weiter, denn hier ist ein wichtiger Unterschied zu Ohalas Modell. In Lindbloms Modell sind - wie wir oben gesehen haben - hypoartikulierte Formen oft von Lautwandel betroffen, wenn sie dekontextualisiert werden. Hypoartikulierte Formen kommen oft in häufigen Wörtern vor (wie Funktionswörtern). Daher ist eine Vorhersage von Lindbloms, jedoch nicht Ohalas Modell, dass im Allgemeinen häufigere Wörter eher als seltenere Wörter von Lautwandel betroffen sind.

#### 5. Unterschied zu Ohalas Modell: 2. Der Sprecher

-----

Hier sind 3.6 und 3.7 von lindblom95.rivling.pdf wichtig. Hier wird wieder betont, dass zwei Faktoren den Bestandteil eines Lautinventars prägen: 1. der Sprecher gibt sich so wenig Mühe wie möglich (= Hypoartikulation), um Laute distinktiv (= Hyperartikulation) für den Hörer zu produzieren. Insofern haben viele Sprachen der Welt [i, e, a, o, u], weil diese perzeptiv distinktiv sind. Man könnte zwar diese Vokale für den Hörer noch distinktiver machen (durch sekundäre Artikulation, Nasalisierung usw.) aber im allgemeinen geschieht das nicht, weil der Sprecher sich mehr Mühe geben würde, als für die Hörer-Verständlichkeit benötigt wird - d.h. Lautinventare von

Sprachen sind laut dieser Theorie meistens genügend distinktiv. Diese zwei Kräfte (sowenig Mühe für den Sprecher wie möglich, so distinktiv wie möglich für den Hörer) prägen auch Lautwandelprozesse - d.h. der Sprecher trägt viel mehr zu dem Lautwandel-Ursprung in Lindbloms Modell bei als in Ohalas. Hier weicht Lindblom auch von Ohala ab, denn artikulatorische Mühe spielt in Lindbloms Modell eine große Rolle - während Ohala die Existenz artikulatorischer Mühe generell bezweifelt.

#### 6. Weitere Unterschiede im Vgl. zu Ohalas Modell: Die Verbreitung des Lautwandels

-----

Ohala trennt sehr scharf zwischen dem Ursprung des Lautwandels (den Bedingungen, die zu Lautwandel führen können) und dessen Verbreitung. Er befasst sich nur mit dem Ursprung. Lindblom versucht, die beiden Teile zu verbinden, d.h. für Lindblom können soziale Faktoren, die zu der Verbreitung des Lautwandels führen, nicht von dessen Ursprung getrennt werden. NB [1] S. 245: 'it may be unnecessary to limit the phonetic contribution to sound change to the initiation stage'.

Ob Variabilität sich in Lautwandel umsetzt, hängt laut Lindblom davon ab, wie der potentielle Lautwandel evaluiert wird. Der Hörer evaluiert den potentiellen Lautwandel auf eine dreifache Weise (siehe lindblom98.pdf, und lindblom95.rivling.pdf S. 28)

(a) articulatory ease (wie aufwendig der Laut ist in der Produktion?)

(b) perceptual adequacy (wie distinktiv der Laut ist für den Hörer?)

(c) social value (ist der Laut sozial akzeptierbar?)

Ein Beispiel für (c): ein /t/ in 'water' in einer südenenglischen Standardaussprache wird selten mit einem Glottalverschluss ersetzt, da dies ein Merkmal einer sozialen niedrigen Klasse (London Cockney) ist.

#### 7. Inwiefern ist Lautwandel evolutionär im biologischen Sinne

-----

Für dieses siehe vor allem Ohalas Paper [4]. Das wichtigste aus [4] ist wie folgt: Ohala differenziert zwischen zwei Formen der Variation in Biologie. Kurzfristig (short-term): das ist die Anpassung von einem Tier an die Umgebung - dies ist laut Ohala die Form der Adaption, die in Lindbloms Paper gemeint wird. Dann gibt es die zweite Sorte der biologischen Variation, die wegen der Änderung von Genen vorkommt, die zu einer Differenzierung von Spezies führt.

Für Ohala ist Lautwandel teilweise ähnlich zu der zweiten Sorte der biologischen Variation, da diese zweite Sorte nicht teleologisch ist (folgt keinem Ziel). Ebenso gibt es folgende Ähnlichkeiten. In biologischer Adaptation führen kurzfristige Anpassungen an die Umgebung zu kategorialen Unterschieden - wie zwischen Okapis und Giraffen. Lautwandel ist auch kategorial (eine Änderung von z.B. /s/ zu /ʃ/ usw.) und entsteht kumulativ aus der ständigen Anpassung von einem Laut an Kontexte.

Jedoch ist Lautwandel laut Ohala doch anders, weil es sich bei Lautwandel *nicht* um einen Wettbewerb zwischen Ausspracheformen handelt (in der Biologie hingegen überlebt die Spezies, die sich am besten an die Umgebung anpasst). D.h. es gibt keinen 'optimization through competition of languages'. Und: 'like scribal errors, there is no adaptive value to such variations...there has been no detectable improvement in the communicative capacity of speech.'

## # 11. Exemplar theories of speech and abstraction

pierrehumbert03b.pdf

=====

Es handelt sich in diesem Paper um einen Überblick zur Exemplartheorie (exemplar theory oder auch episodic models of speech) und vor allem wie Wörter, Phonologie, und die Phonetik zusammenpassen.

Hier werden zwei Standpunkte zusammengebracht.

(a.) Auf der einen Seite könnte es sein, dass die Phonologie so produktiv ist, dass durch ein phonologisches Regelwerk die Aussprache aller Wörter im Lexikon vorhergesagt werden könnte.

(b.) Auf der anderen Seite könnte es auch sein, dass es gar kein phonologisches Regelwerk gibt, und dass statt dessen die Aussprachen alle Wörter einfach unabhängig voneinander aufgelistet und gelernt werden.

Der Standpunkt der generativen Phonologie ist abgesehen von Ausnahmen (a.).

Der Standpunkt der hier geschilderten Exemplartheorie ist, dass ein Kompromiss zwischen (a.) und (b.) gefunden werden muss.

Zunächst einmal sind zwei Sachen in der Exemplartheorie wichtig:

i. Laut der Exemplartheorie speichert man im Lexikon sehr viele Details wahrgenommener Sprachsignale - viel mehr Details als früher (oder in der *generativen* Phonologie) vermutet wurden. Der Ansatz der generativen Phonologie dagegen ist, dass redundante Details gar nicht im Lexikon gespeichert werden (vor allem keine Sprechermerkmale oder sonstige Details), sondern nur die minimalen Merkmale, um zwischen Wortbedeutungen zu unterscheiden. (z.B. speichert man nur die minimalen Unterschiede, damit Wörter wie 'Tank', 'Dank' voneinander differenziert werden können).

ii. Laut der Exemplartheorie sind Wörter und auch die Phonologie als statistische Verallgemeinerung im Gedächtnis gespeicherter Sprachsignale zu verstehen. Ein Beispiel aus der Statistik ist nützlich, um das zu erklären: wenn ich 10 Würfel zusammen werfe, und jedes Mal den Mittelwert berechne, bekomme ich lauter Stichproben wie 3.3, 2.9, 4.3 usw. Diese Stichproben sind analog zu den im Gedächtnis gespeicherten Sprachsignalen.

Man kann aber die Verteilung durch ein Gaussglocke mit zwei Parametern (Mittelwert, Standardabweichung) modellieren. Diese Abstraktion, mit der man die ganzen Stichproben durch eine Handvoll Parameter statistisch modelliert, ist analog zu einer phonologischen Klasse oder auch einer Wortklasse. Daher liest man oft in der Exemplartheorie: 'phonological categories are generalisations over remembered signals' usw.

Mehr Details zu i. und ii. aus dem Paper meistens bis S. 193. sind unten.

Das Paper ab S. 193 kann zum großen Teil überflogen/weggelassen werden.

### 1. Die Notwendigkeit einer Phonologie (S. 178)

+++++

Es muss eine generative Phonologie geben (S. 178). Denn L1-Sprecher sind imstande, neue Wörter zu bilden, die den Regeln der Aussprache folgen. Z.B., wenn ich ein neues Wort für ein Produkt wie ein Waschpulver erfinde, genannt 'Sudsil', dann folgt dieses Wort verschiedenen Regeln:

ich bilde meistens keinen Namen, der gegen die Phonotaktik verstößt. Z.B. 'Fnudsil' käme eher nicht auf den Markt, weil /fn/ im Anlaut im Deutschen nicht vorkommt.

/s/ in Sudsil wird in der Standardsprache meistens [z] sein; der silbenfinale Konsonant in 'sud' wird wegen Auslautverhärtung usw. meistens [t] sein. Dies zeigt, dass es Regeln geben muss (also ein phonologisches Regelwert), die auf Logatome oder neue Wörter angewandt werden.

Phonologische Einheiten, d.h. *phonologische Abstraktion* ist notwendig, weil dadurch auch die Perzeption der Sprache schneller wird. Denn wenn sich Wörter im Lexikon in abstrakte phonologische Einheiten zerlegen lassen, dann kann man im Lexikon Wörter schneller finden, die zu dem wahrgenommenen Signal passen (z.B. wenn ich erkenne, dass ein Wort mit einem /s/-ähnlichen Laut beginnt, dann brauche ich nicht in Wörtern zu suchen, die mit /n/-beginnen. Aber das setzt eine phonologische Zerlegung von Wörtern voraus). Phonologische Abstraktion bedeutet auch Flexibilität: man kann schnell und problemlos Logatome (also neue Wörter) durch die Kombinatorik der phonologischen Einheiten bilden (siehe oben), auch wenn solche Reihenfolgen von Phonemen noch niemals vorher wahrgenommen worden sind.

Im Allgemeinen S. 194: 'the productivity of phonology indicates that humans have internalized an abstract system for making complex forms from simpler parts'.

### 2. Wörter in der Exemplartheorie

+++++

i. Ein Wort im Lexikon setzt sich aus allen auditiven Eigenschaften des wahrgenommenen Wortes inkl. der phonetischen Details und Sprecherinformationen zusammen; d.h. jedes Mal, wenn ein Wort gesprochen wird, hinterlässt es eine kontinuierliche, nicht segmentierte Spur des akustischen Signals (= *parametric phonetics*) im Gedächtnis des Hörers.

Dadurch (siehe S. 187) enthalten Wörter viele gespeicherte Informationen, darunter auch Informationen zu den Stimmen der Personen, die die Wörter gesprochen haben, und viele andere phonetische Details. Z.B. gibt es feine Unterschiede in dem ersten /t/ (s 181) von 'capitalistic' und 'militaristic', weil ein Tap in dem Stamm von 'capital' aber nicht von 'military' vorkommt.

Diese Idee, dass sehr viele Informationen im Lexikon pro Wort gespeichert werden, steht im Gegensatz zur generativen Phonologie: in der generativen Phonologie werden

alle *redundanten Details* des Wortes aus dem Lexikon entfernt, und durch Regeln, die auf mehrere Wörter gleichzeitig angewandt werden, postlexikalisch hinzugefügt. Es kann aber keine Regel geben, die ein Tap in 'capitalistic', aber nicht in 'militaristic' vorhersagt - gerade weil diese Infos *wortspezifisch* (und daher *gelernt*) sind, und daher mit dem Wort (laut der Exemplartheorie, die Pierrehumbert hier schildert) gespeichert werden müssen.

Wortspezifische Details entstehen auch aufgrund der Nachbarschaftsdichte und Häufigkeiten (S. 188-191):

Nachbarschaftsdichte (neighbourhood density).

Ein Wort wie 'pin' im Englischen hat eine hohe Nachbarschaftsdichte, weil es viele andere Wörter im Lexikon gibt, auch wenn man nur ein Phonem davon austauscht (bin, fin, sin, tin, win, kin).

Ein Wort wie 'help' hat eine kleinere Nachbarschaftsdichte - es gibt kaum andere Wörter, wenn man /h/ oder /e/ oder /l/ austauscht.

Worthäufigkeit. Dies ist, wie oft ein Wort in der gesprochenen Sprache vorkommt/ verwendet wird. Funktionswörter haben eine hohe Worthäufigkeit.

Die daraus entstehenden wortspezifischen phonetischen Details sind wie folgt:

Hörer brauchen mehr Zeit, um Wörter zu erkennen,

(a) die eine hohe Nachbarschaftsdichte haben, weil sie leicht mit anderen Wörtern verwechselt werden, und

(b) wenn sie nicht so oft vorkommen, weil seltene Wörter auch meistens nicht durch den Kontext vorhersagbar sind.

Sprecher produzieren oft solche Wörter (= hard words) mit etwas mehr Klarheit (weil sie für den Hörer verwechselbarer sind, und weil sie nicht so leicht vorhergesagt werden können). Dies zeigt, dass man die phonologische Information eines Wortes von der Erfahrung damit (wie oft man ein Wort gehört/produziert hat) nicht trennen kann.

### 3. Phonetische Kategorien in der Exemplartheorie

+++++

Die Einheiten der phonetischen Enkodierung entstehen im Gedächtnis oft an den Stellen, wo sich die Spuren gespeicherter Wortexemplare in einem multidimensionalen akustischen Raum überlappen. Phonetische Einheiten sind also eine Abstraktion der (akustisch-phonetischen) Details in 2. Es ist wichtig, hier zu verstehen, dass phonetische Einheiten aufgrund der gespeicherten Spuren entstehen (insofern ist die Beziehung zwischen phonetischen Einheiten und Sprachsignalen etwas anders pro Person, da ja keine zwei Personen die selben Wortspuren gespeichert haben werden). Daher 'they can be viewed as peaks in the total phonetic distribution of the language'. *Peaks* ist hier wichtig, weil das bedeutet, dass eine phonetische Einheit eine statistische Verallgemeinerung der damit verbundenen Sprachsignale ist; d.h., eine phonetische Einheit wie [n] entsteht, weil sich eine Wolke von Spuren im Gedächtnis bildet, und diese Wolke ändert sich jedes Mal, wenn ein weiteres [n] vorkommt. Diese Wolke wird dann parametrisiert - ähnlich z.B. wie man in der Statistik eine Verteilung von Punkten durch eine Gaussglocke (Mittelwert, Standardabweichung) parametrisieren kann (siehe oben). Die Abstraktion einer phonetischen Einheit entsteht gerade deswegen, weil sie eine statistische Verallgemeinerung über eine dichte Wolke von Signalen ist.

Diese phonetischen Kategorien sind enger als Phoneme. z.B.

das /t/ in 'Tau' und der /t/ in 'Stau' sind phonetisch unterschiedlich (aspiriert in 'Tau' / nicht aspiriert in 'Stau'). Daher gibt es eventuell getrennte phonetische Einheiten für diese beiden /t/s, weil die auditiven Spuren sich für diese ganz unterschiedlichen /t/s kaum überschneiden werden. Daher: 'the phonetic categories are less abstract than phonemes...they can be viewed as positional allophones.'

Ein wichtiger Punkt ist der Folgende: sowohl phonetische Kategorien, also auch die Phonologie und Wörter sind *abstrakte* Kategorien, die aufgrund der *Erfahrung* mit der Sprache entstehen. Die *Erfahrung* = welche Wörter eine Person je in seinem Leben gehört hat. Das erklärt dann auch, weshalb die Beziehung zwischen Phonologie und Sprachsignalen zum Teil auch stochastisch/willkürlich ist. Sie erscheint willkürlich, weil keine zwei Personen die selben Sprecherfahrungen gemacht haben werden. Diese willkürliche Beziehung zwischen Phonologie und Sprachsignalen erklärt auch (a) weshalb Aussprachen phonetisch so detailliert sind - weshalb also eine Aussprache so viele willkürliche, idiosynkratische phonetische Merkmale hat. Das erklärt auch (b) weshalb zwei Sprachen mit ähnlichen phonologischen Kontrasten sich immer etwas phonetisch unterscheiden. z.B. haben englisch und deutsch beide /l/, /i:/ und /f/ Phoneme - aber die Phonetik von dt. 'lief' und engl. 'leaf' (Englisch <Blatt>) ist keineswegs identisch (daher im Paper: 'there is no known case of a phoneme that has exactly the same phonetics in two different languages').

#### 4. Phonologie (phonological grammar).

+++++

Die Phonologie definiert die Wahrscheinlichkeit, dass gewisse Lautreihenfolgen ein Wort bilden können.

Z.B. ein Logatom wie /tmsɔ/ hat im Deutschen und vielen anderen Sprachen eventuell eine Wahrscheinlichkeit von 0, weil solche Reihenfolgen für die meisten Hörer nie vorgekommen sind.

Ein Logatom wie /ʃno:/ hat eine Wahrscheinlichkeit von 1, da Wörter mit der Reihenfolge /ʃn/ (z.B. 'Schnee', 'Schnecke' usw. ) tatsächlich vorkommen.

Ein Logatom wie /gmɑu/ hat eine Wahrscheinlichkeit irgendwo dazwischen, da, obwohl es im Standarddeutschen keine Wörter mit initialem /gm/ gibt, /gm/ in reduziertem 'gemeint' /gmaint/ usw. vorkommen kann (vor allem im bairischen Dialekt).

Das neue hier ist also: Phonologie bildet mögliche Wörter aufgrund von phonotaktischen *Wahrscheinlichkeiten*. Dies steht im Widerspruch zu der früheren generativen Phonologie (von Chomsky & Halle, 1968, und anderen). Hier gab es Regeln: entweder ist ein Reihenfolge *phonotaktisch legal* wie /ʃn/ *oder nicht* wie /tms/. Dieses Regelwerk der generativen Phonologie konnte aber diese Zwischenfälle wie /gm/ nicht erklären.

Die Evidenzen, dass die Phonotaktik probabilistisch ist, werden auf S. 191 erläutert. Hörer neigen dazu, im Lärm phonotaktische Reihenfolgen, die nicht so häufig

vorkommen, als häufigere wahrzunehmen. Dies war das Ergebnis von Hay, Beckman & Pierrehumbert (2000). Man verwechselt /nf/ (Senf) öfters mit /nʃ/ (Mensch) und nicht umgekehrt, weil /nʃ/ so selten vorkommt. Dies ist wieder ein Hinweis dafür, dass die Phonetik nicht als ein binäres Regelwerk gespeichert wird, sondern eher zusammen mit Wahrscheinlichkeiten.

#### 5. Spracherwerb (S. 185)

+++++

Die Fähigkeit zur phonologischen Abstraktion entsteht im Laufe des Spracherwerbs - es passiert nicht auf einmal, sondern ist zum Teil eine Funktion der Vergrößerung des mentalen Lexikons: denn je mehr verschiedene Wortspuren wahrgenommen und im Gedächtnis gespeichert werden, desto wahrscheinlicher entstehen Knoten/Dichtigkeiten, die die Voraussetzung für die Phonologisierung (= Fähigkeit zur phonologischen Abstraktion/Entstehung phonologischer Einheiten) sind.

Daher auch S. 187: 'infants acquire their first words before they abstract phonemes'.

#### cutler10.labphon.pdf

+++++

Die Argumente in diesem Paper untermauern die Notwendigkeit der phonologischen Abstraktion (prelexical information) diesmal anhand von psycholinguistischen Experimenten.

Es werden drei alternative Modelle vorgestellt:

1. Ein episodisches Modell ohne Phonologie (S. 93). In einem solchen reinen episodischen Modell (z.B. das von Goldinger) werden alle Spuren von Wörtern zusammen mit allen Merkmalen des Sprechers gespeichert, wie oben erklärt, aber es gibt *keine* Phonologie. Wenn ein Hörer ein Sprachsignal wahrnimmt, dann werden akustisch ähnliche Spuren im Lexikon aktiviert. Das Wort wird von Hörer erkannt, dessen Spuren am meisten aktiviert werden (z.B. aktiviert ein Sprachsignal von Haus recht viele im Gedächtnis damit verknüpfte Wortspuren, oder zumindest mehr Spuren als in anderen, damit phonetisch unverwandten Wörtern).

2. Ein phonologisches Modell ohne Informationen zu dem Sprecher.

Dies ist der Ansatz der generativen Phonologie. Laut der generativen Phonologie werden Sprechereigenschaften nicht im Lexikon aufgenommen, wie oben erklärt. Um ein Wort zu erkennen, werden daher die Sprecherinformationen zuerst weggefiltert. Dies setzt voraus, dass Sprecher- und phonologische Eigenschaften getrennt gespeichert werden - also dass Sprechereigenschaften nicht im Lexikon aufgenommen werden. Es gibt zwar einige empirische Beweise, dass diese scharfe Trennung zwischen Phonologie und Sprecherinformation richtig sein könnte (S. 92). Z.B. gibt es Sprecher mit Aphasie, die nicht die Wörter, sondern nur die Identität des Sprechers erkennen; und anders herum Patienten, die die Phonologie, aber nicht die Sprecheridentität erkennen.

Jedoch kann ein kognitives Modell, in dem Informationen zum Sprecher und zur Phonologie komplett unabhängig voneinander gespeichert werden, auch nicht richtig sein, und zwar wegen Evidenzen, dass Sprecherstimmen die Wahrnehmung linguistischer Informationen (S. 92/93) beeinflussen. Z.B. identifiziert man Wörter schneller/genauer von einem Sprecher, den man schon einmal wahrgenommen hat,

eventuell gerade weil seine Merkmale gespeichert werden. Zusätzlich wird ein phonologischer Kontrast in einer fremden Sprache schneller gelernt, wenn man den Kontrast mehrerer Muttersprachler wahrgenommen hat: also ist das Erlernen eines phonologischen Kontrastes in einer anderen Sprache davon abhängig, welche (und wieviele) Muttersprachler man beim Produzieren dieses Kontrastes gehört hat.

Daher S. 93:

'there is thus strong evidence that talker-specific information plays a role in speech perception. An extreme abstractionist view, in which talker-specific information is discarded during lexical access, is therefore untenable.'

3. In diesem Paper wird also betont, dass weder 1 (ein Lexikon ohne Phonologie) noch 2 (Lexikon ohne Sprechermerkmale) richtig sein kann. Es geht also darum, festzustellen, wie genau Sprechermerkmale mit phonologischen Merkmalen in der Worterkennung anhand von lexically-guided (perceptual) learning miteinander interagieren - wie in 4. zusammengefasst:

#### 4. Lexically-guided learning (S 94)

+++++

Diese Experimente zeigen, dass die kategoriale Grenze zwischen zwei Lauten flexibel an die Eigenschaften des Sprechers angepasst werden. Das Experiment hat zwei Komponenten: Training und Test.

#### Training

-----

Es handelt sich um Wörter mit finalelem /s/ und andere Wörter mit finalelem /f/ *gesprochen vom selben Sprecher*

Das finale /f/ wird resynthetisiert, sodass es ambig zwischen /f/ und /s/ ist. Z.B. wäre das /f/ vom deutschen 'Wurf' ambig, wenn das /f/ mit Merkmalen von /s/ resynthetisiert wird (NB: Wörter wurden immer so gewählt, dass durch diese ambigen akustischen Merkmale keine Minimalpaare gebildet wurden: z.B. ein Wort wie Wurf, weil es kein Wort 'Wurs' gibt).

Diese Wörter mit finalelem /s/ und Wörter mit ambigem /f-s/ wurden Hörern präsentiert. Das ist die Trainingsphase.

#### Test

-----

Danach wurde den Hörern ein Kontinuum zwischen /s/ und /f/ *mit Merkmalen desselben Sprechers* (also man hört im Kontinuum immer dieselbe Sprecherstimme) präsentiert. Hörer mussten pro Stimulus sagen: war das ein /s/ oder /f/? (= forced choice experiment).

Die Ergebnisse zeigen, dass Hörer viel mehr von dem Kontinuum mit /f/ identifizieren, nachdem ihnen in der Trainingsphase Wörter mit eindeutigem /s/ aber ambigem /f-s/ präsentiert wurde. Der Grund: die Wahrscheinlichkeit, dass ein ambiger Laut zwischen /f-s/ in der Testphase ein /f/ sein könnte, ist viel größer, wenn die Hörer vorher gelernt hatten, dass dieser Sprecher seine wortfinalen /f/s auf eine ambige Weise produziert.

Die Ergebnisse funktionieren genauso in die andere Richtung. Wenn im Vortest die Wörter mit finalelem /s/ ambig sind (zB 'las' wird als la[s-f] synthetisiert) und die Wörter mit finalelem /f/ ein eindeutiges /f/ haben, dann ist die Wahrscheinlichkeit im forced-choice Test größer, dass Hörer denselben ambigen Laut aus dem /s-f/ Kontinuum als /s/ interpretieren.

Diese Experimente zeigen also, dass die Phonologie (ob ich ein /s/ oder ein /f/ wahrnehme) *flexibel an die Sprechereigenschaften angepasst wird*.

Im Abschnitt 3. 'stability of learning' gibt es zusätzlich die folgenden nützlichen Infos:

- inwiefern dieser Effekt über die Zeit stabil ist: Eisner & McQueen (2006) zeigen, dass *perceptual learning* auch funktioniert, wenn 12 Stunden zwischen Training und Test vorliegen.

- inwiefern (S. 98) der Effekt auf andere Wörter übertragen wird: Es handelt sich um das gleiche Experiment, aber diesmal mussten Hörer aufgrund von einem präsentierten [do:?] zwischen DOOS und DOOF auf dem Bildschirm wählen. Es gab mehr DOOF Antworten, wenn Hörern in der Trainingsphase /s/-Wörter mit eindeutigem finalelem /s/ und /f/-Wörter mit ambigen finalelem /s-f/ präsentiert wurden. Das wichtigste aber: DOOS oder DOOF kamen in der Trainingsphase NICHT vor. Daher wird diese Änderung in der Grenze zwischen /s/ und /f/ (die dazu führt, wie ein Hörer ein Wort mit einem finalelem ambigen Stimulus erkennt) *auf neue Wörter übertragen (= generalisiert)*.

Ohne eine Phonologie könnte ein solches Ergebnis nicht erklärt werden. D.h. der Effekt von perceptual learning breitet sich auf Wörter wie DOOF/DOOS mit einer ähnlichen Phonologie aus. Daher die Schlussfolgerung: 'These results therefore suggest that the perceptual adjustment induced during training affected a prelexical stage of processing, allowing learning to transfer to other words in the lexicon'.

#### 5. Simulations with an episodic model

+++++

Es ist wirklich keineswegs wichtig, das computationelle Experiment zu erklären, sondern eher einen weiteren Grund zu erläutern, weshalb diese Ergebnisse in 4. nicht durch ein episodisches Modell (1.) erklärt werden können.

Und zwar wie folgt (s. 100-101):

Im episodischen Modell - wie oben unter 1. beschrieben - verursacht ein wahrgenommenes Sprachsignal eine Aktivierung aller Spuren im Lexikon, die ähnlich wie das wahrgenommene Signal sind. Es gibt selbstverständlich sehr viele Spuren pro Wort, gerade weil man Wörter so oft gehört hat. Ambige Stimuli würden dann wie ein Tropfen auf dem heißen Stein wirken, d.h. diese ambigen Stimuli wären so sehr in der Minderheit im Vgl. zu den bereits gespeicherten Spuren, sodass die Trainingsphase mit z.B. eindeutigen /s/ und ambigen /f/ keine Wirkung auf die Klassifizierung in der Testphase haben könnte.

Das hätte zur Folge, dass ein Wort mit ambigen Stimulus wie /do?/ genauso viele DOOF wie DOOS Wörter aktivieren müsste - aber das ist nicht was das Ergebnis der Experimente in 4. zeigt.

Der Rest vom Paper 102-106 mit der Modellierung kann überflogen/weggelassen werden.

## # 12. Sound change and word frequency

hay15.cognition.pdf

+++++

Es handelt sich hier um eine akustische Analyse der neuseeländisch-englischen Aussprache über einem Zeitraum von mehr als 100 Jahren. Die Autoren testeten die Hypothese, ob sich Lautwandel zuerst in häufig vorkommenden Wörtern stattfindet. Sie finden aber genau das Gegenteil: Lautwandel wird eher durch seltene Wörter geleitet.

Sie differenzieren zwischen zwei Sorten von Lautwandel: das ist nichts neues - das kam auch bei den Junggrammatikern (Neogrammarians; eine Gruppe von Sprachwissenschaftlern des letzten Drittels des 19. und noch des beginnenden 20. Jh) vor.

Regulärer Lautwandel (das ist Lautwandel, von dem bislang in diesem Seminar die Rede war) entsteht aus phonetischen Gründen, und betrifft meistens alle Wörter, die von dem Kontext betroffen werden. Z.B. in fast allen Wörtern im Französischen entwickelten sich nasalisierte Vokale vor Nasalkonsonanten (main < manus usw.). Das hat offenbar phonetische Gründe (wegen der Koartikulation usw.). Laut den Junggrammatikern verbreitet sich dieser reguläre Lautwandel gleichzeitig in *allen* davon betroffenen Wörtern (es ist also laut der Junggrammatiker keineswegs der Fall, dass sich häufige Wörter zuerst ändern).

*Analogischer* Wandel betrifft meistens nur individuelle Wörter und hat oft keine phonetischen Gründe. Z.B. im Englischen haben manche Personen in der Vergangenheitsform von 'dive' die Form 'dove' ('tauchte') (anstatt des üblichen 'dived') *analog* zu 'rode' von 'ride'. Offenbar hat dieser Wechsel 'dived' -> 'dove' nichts mit synchroner Variabilität zu tun, sondern eher weil die Alternation in 'ride'/'rode' übertragen worden ist. Das kommt nur in diesem Wort vor, und ist auch keine kontinuierliche Änderung ('dived' entwickelt sich phonetisch nicht zu 'dove' durch eine allmähliche Änderung von /ai/ nach /ou/, sondern ändert sich abrupt. Einmal sagt man 'dived', dann 'dove'.

Dieses Paper behandelt den regulären (kontinuierlichen, phonetischen) Wandel und nicht den analogischen Wandel.

Die Hauptfrage ist hier also, ob Lautwandel gleichzeitig alle Wörter betrifft, wie es die Junggrammatiker behaupteten, oder ob Lautwandel zuerst in häufigen Wörtern vorkommt.

Das letzte wäre z.B. eine Vorhersage von Lindbloms Modell, da häufig vorkommende Wörter so oft hypoartikuliert sind.

Das neue und interessante an diesem Paper ist die Evaluierung dieser Hypothese anhand von einem Archiv gesprochener Sprache, ONZE (S. 84).

## 2. lexical frequency & regular sound change

Exemplartheorie sagt auch voraus, dass Lautwandel zuerst in häufig vorkommenden Wörtern stattfinden müsste. Das ist

(a) weil ein Wort eine statistische Verallgemeinerung über dessen gespeicherten Signalen ist; und

(b) weil diese Verteilung jedes Mal upgedatet wird, wenn das Wort wahrgenommen wird.

Wenn sich ein Lautwandel durch die Gemeinschaft verbreitet, dann werden viel mehr Exemplare mit diesem Lautwandel in häufig vorkommenden Wörtern gespeichert, gerade weil sie so oft vorkommen. Insofern werden die Verteilungen schneller in häufig als in selten vorkommenden Wörtern von dem Lautwandel-im-Fortschritt beeinflusst.

Jedoch behaupten andere (Labov, 1994; Kiparsky, 2014), dass Lautwandel nicht zuerst in häufigen Wörtern vorkommt.

Keiner hat bislang behauptet (S. 85), dass Lautwandel zuerst in selten vorkommenden Wörtern stattfindet.

S.84: Es ist nicht einfach diese Theorie (dass Lautwandel schneller in häufig vorkommenden Wörtern stattfindet) zu testen. Es reicht nicht aus, zu zeigen, dass Verschleifungen usw. in häufig vorkommenden Wörtern stattfinden. Man muss auch zeigen, dass die Häufigkeit mit dem Lautwandel *interagiert*, also dass die Aussprache von häufigen und seltenen Wörtern sich am meisten während eines Lautwandels-im-Fortschritt unterscheiden.

Es hat in den letzten 50-100 Jahren eine sehr interessante Vokalverschiebung im Neuseeland-Englischen (NZE) gegeben:  $/æ \rightarrow \varepsilon \rightarrow \text{I} \rightarrow \text{ə}/$ . Das ist eine Kettenverschiebung, denn die Verschiebung betrifft mehrere Vokale gleichzeitig. HAD hört sich in NZE wie HEAD an; HEAD ist wie HID; und HID ist  $/həd/$ . Siehe Folie nze.ppt (die auch verwendet werden kann) mit Sound-Beispielen in  $/vdata/Seminar/prosodie/lit$ . Es gibt zwei Arten einer solchen Kettenverschiebung: Push chain: ein Vokal schiebt einen anderen zu einer anderen Position; drag-chain: wenn ein Vokal sich verschiebt, dann gibt es sozusagen ein Vakuum, wo sich dieser Vokal früher befand. Ein drag-chain ist, wenn ein anderer Vokal hier in das Vakuum hineingezogen wird (= die Stelle belegt, die früher vom anderen Vokal belegt wurde). Offenbar ist NZE irgendeine Art von Kettenverschiebung. Eventuell push-chain ausgelöst durch  $/æ \rightarrow \varepsilon/$

Es handelt sich bei Kettenverschiebungen um das Aufrechterhalten der Wortkontraste: es ändert sich also nur die Phonetik, d.h. Kontraste gehen nicht verloren wie in Auslautverhärtung, in der Wörter wie Rat/Rad neutralisiert worden sind (= haben die selbe Aussprache).

Die Hypothese (S. 85, Spalte 2) ist wie folgt. Wenn zwei Vokale sich wegen Lautwandel nähern, gibt es einen Bereich der Überlappung, in der man z.B. nicht mehr weiß, ob der Sprecher *had* oder *head* meinte. Die Hypothese ist: seltene Wörter kommen in dieser zweideutigen Region nicht vor, weil es sowieso schwieriger sein wird, seltene Wörter zu verstehen (weil der Hörer weniger imstande ist, die Semantik/den Kontext bei seltenen Wörtern einzusetzen, um das Wort zu verstehen). Daher kommen seltene Wörter nicht so sehr in dieser ambigen Region vor. Die Hypothese der Autoren sagt statt dessen, dass

Lautwandel eher häufig vorkommende Wörter betrifft, die viel mehr in dieser ambigen Region auftreten. (Das ist die Hypothese, die aber dann NICHT bestätigt wird).

Methode eher ganz kurz zusammenfassen. Hier soll man vielleicht die Mengen beachten: 2741 verschiedene Wörter und 549 verschiedene Sprecher mit Geburtsdatum zwischen 1851 (! ja, 1851!) und 1987.

F1 und F2 Messungen vor allem in betonten Vokalen.

Ergebnisse: es reicht aus, Fig. 1 zu präsentieren. Solide Linien = Vokale in seltenen Wörtern, gestrichelt = häufig vorkommende Wörter. Also sind die Ergebnisse in die entgegengesetzte Richtung zur Vorhersage der Hypothese! D.h. der Wandel ist größer für seltene als für häufig vorkommende Wörter. Fig. 2 kann man weglassen. Das dupliziert nur die Informationen in Fig. 1.

### Diskussion

'For all three vowels, low frequency words lead the way.'

Was könnte also die Erklärung dafür sein? Es ist kompliziert (wie immer), ca. wie folgt:

Es kommen zwei Sachen zusammen: Es genügt, hierzu Fig. 3 zu präsentieren. (NB BET steht für die Wortklasse (lexical set) mit /ε/ Vokalen (bet, said, met, head, fed...) nicht nur für den Vokal im Wort bet; BET ist nur das keyword für dieses lexical set).

(a) Zuerst muss man erklären, weshalb BET durch BAT (=push chain) weggeschoben wird. Das funktioniert so: Wenn BET-Wörter in der Region der Überlappung produziert werden - d.h. wo sich die Ellipsen überlappen - nimmt der Hörer sie zwar korrekt als BET wahr, aber er ergänzt seine kognitive BET-Verteilung nicht damit. Das ist so, weil seine kognitive BET-Verteilung (oder generell Vokalverteilungen) nur mit klaren, eindeutigen Exemplaren von BET ergänzt werden. Andere wahrgenommene Vokale in der Region der Überlappung speichert man erst gar nicht im Gedächtnis.

Wenn daher die BET-Stichproben, die sich im überlappenden Teil befinden, nicht in die kognitive BET-Verteilung aufgenommen werden, dann schiebt sich der Mittelwert von BET weg von BAT (weil der überlappende Teil der BET-Verteilung weggeschnitten wird). Ähnlich wie: wenn ich 100 Stichproben habe mit Mittelwert von 3 (analog zu Mittelwert von BET) und ich werfe alle Stichproben unter 1.5 weg (alle im überlappenden Teil), dann verschiebt sich der Mittelwert der Verteilung nach oben (Richtung 4, weg von BAT). So ähnlich - laut der Erklärung der Autoren - in einem push-chain.

(b) Meistens werden seltene Wörter nicht in dieser überlappenden Region produziert. Der Grund: sie sind für den Hörer nicht aus dem Kontext so gut vorhersagbar, gerade, weil sie kaum vorkommen. Das ist nicht so für häufig vorkommende BET-Wörter, die oft in dieser überlappenden Region auftauchen können. Das ist möglich - gerade weil der Hörer nicht so stark auf das Signal angewiesen ist, um häufig vorkommende Wörter zu erkennen.

Das bedeutet also: *seltene* BET-Wörter verschieben sich schneller weg von BAT, weil sie Bestandteil der Verteilung sind, die nicht mit BAT überlappt. Analog zur Statistik: wir sagten, der Mittelwert von 3 verschiebt sich in Richtung 4 wenn wir alles unter 1.5

wegschneiden (unter 1.5 ist analog zum überlappenden Teil in Fig. 3). Seltene Wörter sind also Bestandteil dieser Stichproben, die sich von 3 auf 4 ändern, *weil sie sich nicht im überlappenden Teil (unter 1.5) befinden.*

Daher ändern sich seltene Wörter in einem push-chain zuerst.

linn14.laphon.pdf

+++++

Im Englischen werden silbenfinale /l/s velarisiert, d.h. mit Zungendorsumhebung als [ɫ] produziert. Sie ähneln daher in der Produktion und in der Perzeption [u, o], weil diese auch mit Zungendorsumhebung gegen den weichen Gaumen produziert werden. Daher hört sich engl. 'feel' oft wie ein Diphthong an, weil die Bewegung von /il/ wie ein Diphthong ist - von einem vorderen /i/ zu dem velarisierten, rückverlagerten Teil vom /l/ (Daher haben auch die nordenglische Städte *Keele* und *Kiel* ganz andere Aussprachen, weil im deutschen nicht velarisiert wird).

Üblicherweise wird ein /l/ zusätzlich mit einem alveolaren Kontakt produziert. Wenn das /l/ vokalisiert wird, dann fällt dieser alveolare Kontakt oft weg. Daher sind vokalisierte /l/s eher Vokale. In der Londoner-Cockney Aussprache wird häufig vokalisiert. Daher [wo:] für 'wall' usw.

Die Vokalisierung ist mit Lautwandel verbunden. z.B. Französisch 'animaux' (Tiere) war animal(is) -> animals -> [animo]: das 'aux' = [o] kam zustande, weil das finale /l/ im Lateinischen animal stark velarisiert war.

Hier soll geprüft werden, ob die Vokalisierung früher in häufigen Wörtern wie *help, milk* als in seltenen Wörtern wie *whelp, ilk* vorkommt.

Aber noch interessanter ist die Verbindung zu Lautwandel. Wie gesagt, ein velarisiertes /l/ wird oft zu [o, u] wie in 'animaux'. Das bedeutet: der Zungenspitzenkontakt für /l/ fällt weg, *aber die Dorsum-Hebung bleibt*. Es kann also nicht sein, dass der Wandel von velarisiertem /l/ -> [o] allgemein wegen Hypoartikulation zustande kommt, weil man dann auch eine Zungendorsumsenkung (Richtung Schwa) erwarten würde, und das ist nicht der Fall. Insofern müsste man eine Lenisierung (Abschwächung, bis hin zum Wegfall) des alveolaren Kontaktes, bei gleichzeitiger Beibehaltung der Zungendorsumhebung in den Ultraschall-Aufnahmen sehen.

Hier ist diese Frage, ob dieses Muster (Alveolar fällt weg, Zungendorsumhebung bleibt) eher in häufigen als in seltenen Wörtern vorkommt.

Ein weiter treibender synchroner Faktor in diesem Lautwandel (S.12) ist wie folgt: Ein alveolarer Kontakt für /l/ zieht das Zungendorsum nach vorne, wodurch F2 angehoben wird. Daher ist es der Fall, dass, wenn vokalisiert wird (= kein alveolarer Kontakt), dann geht F2 nach unten. Zusätzlich wird diese F2-Senkung noch deutlicher, wenn kein alveolarer Kontakt vorhanden ist, weil dann wird das akustische Signal überwiegend

von der Zungendorsumhebung wie für [o, u] geprägt - und in [u, o] ist F2 niedrig und nah an F1.

Insofern: das akustischen Merkmal für Vokalisierung ist eine Senkung von F2 in Richtung von F1, die noch deutlicher in häufigen Wörtern erscheinen müsste, wenn dieses Muster der Vokalisierung sich zuerst in häufigen Wörtern durchzieht.

Schließlich: es könnte kurz erwähnt werden (S.13), dass die Vokalisierung häufiger vor labialen und vor velaren als vor alveolaren vorkommt. Siehe dazu auch Ergebnisse in Fig. 3 S. 20.

#### Methode

Es handelt sich um Ultraschall-Aufnahmen. Vielleicht Fig. 1 zeigen und wie man die Zunge sieht - vor allem den gleichzeitigen alveolaren Kontakt und die Zungen-Dorsum-Hebung. Vielleicht auch Fig. 2.

Sonst kann die Methode sehr schnell überflogen werden.

#### Ergebnisse

Ebenfalls ganz gezielt auf die Ergebnisse eingehen:

Fig. 3: mehr Vokalisierung vor alveolaren ('The analyses indicate that tongue tip aperture in laterals was significantly smaller') -- aber (Fig. 4): keine Unterschiede in der Zungendorsum-Hebung. Daher Vokalisierung ohne Hypoartikulation des Zungendorsums.

Fig. 5. Treffer! Mehr Vokalisierung in häufigen Wörtern (Fig. 6 eigentlich das gleiche nur aufgeteilt nach Artikulationsstelle)

Dann ab 3.3 S. 25

kann alles weggelassen werden bis auf Figure 7

Diese zeigt:

(a) je mehr Vokalisierung, desto kleiner ist der F2-F1 Unterschied = desto näher ist F2 an F1. Daher haben die Werte unten rechts einen grossen 'tongue tip aperture' = große Vokalisierung und daher ist F2-F1 sehr klein (oft unter 0, weil die Werte sprechernormalisiert wurden; aber das ist nicht so wichtig).

(b) In häufig vorkommenden Wörtern ist der Aperture größer und F2-F1 ist dementsprechend kleiner.

#### Diskussion

(Abb. 8 kann weggelassen werden)

Wiederholt meistens, was schon oben besprochen wurde, aber hier kommt etwas wichtiges und neues hinzu (S. 32):

'For reduction of the lateral alveolar constriction in high-frequency words to serve as the impetus for sound change, it follows that the reduced gesture must have acoustic consequences.'

Das ist ganz wichtig, und etwas kompliziert wie folgt:

Die Beziehung zwischen artikulatorischer Vokalisierung und auditiver Interpretation ist nicht-linear. Das bedeutet, wenn ich zunehmend vokalisierere, geht die Wahrnehmung des Signals nicht progressiv in die Richtung von [o, u]. d.h. wenn Vokalisierung zunehmend vorkommt, höre ich [l, l, l, l, ...] und dann irgendwann kippt die Wahrnehmung plötzlich um auf [u, o]. Warum ist das so? Weil, wenn F2 sich progressiv F1 nähert, diese Formanten ab diesem kritischen Umkipppunkt voneinander nicht mehr differenziert, sondern perzeptiv integriert werden. Und genau zu dem Punkt der perzeptiven Integration von F2 und F1 höre ich nicht mehr [l] sondern [u, o]. Warum ist das wichtig? Weil es zeigt, wie eine kleine, allmähliche kontinuierliche Änderung in der Artikulation *einen plötzlichen Wandel* von einer Kategorie (zB Okapi, [l]) in eine andere (Giraffe, [u, o]) auditiv verursachen kann. Und dieser kategoriale Wandel kommt eventuell zuerst in häufigen als in seltenen Wörtern vor.

zellou14.jop.pdf

+++++

Es handelt sich hier um die koartikulatorisch bedingte Vokalnasalisierung in der amerikanisch-englischen Varietät von Philadelphia-Englisch. Es ist eine 'apparent-time' Untersuchung, d.h. der Lautwandel wird durch einen Vergleich zwischen jüngeren und älteren Personen festgestellt (unter der Annahme, dass der Lautwandel von jüngeren begonnen wird).

Was gefunden wird:

-----

-große Nasalisierung für Leute geb. zwischen 1950 und 1965

-kleine Nasalisierung für Leute geb. zwischen 1965 und 1980

-große Nasalisierung für Leute geb. ab 1980

Was neu ist: Vokalnasalisierung als Lautwandel kann zuerst abnehmen und dann zunehmen.

Man soll aber auf diese Ergebnisse oben nicht näher eingehen (oder ganz und gar weglassen, weil das ist nicht das, was uns hier interessiert... )

Wir wollen uns statt dessen auf Worthäufigkeit (dem Thema des Vortages) konzentrieren. Insofern können alle anderen Teile weggelassen/überflogen werden. Daher aus der Einleitung:

Worthäufigkeit, S. 21

Die Idee ist wie folgt. Häufige Wörter werden hypoartikuliert (siehe H&H Theorie von Lindblom, Thema 10 oben).

Was aber nicht klar ist, ist, ob in hypoartikulierter Sprache mehr koartikuliert wird. Es könnte z.B. auch sein, dass in *hypER*artikulierter Sprache mehr koartikuliert wird, um das Signal zu verdeutlichen (z.B. wenn ich VCV-Koartikulation verdeutliche, dann wird das [a] von Mittag schon im [i] wahrgenommen; auf diese Weise hätte man als Hörer schnelleren Zugriff auf dieses Wort).

Die Sprecherliste kann so zusammengefasst werden, ohne näher auf trend/panel/pseudo-panel (S. 22) einzugehen:

i. 46 Sprecher 'under the age of 25, whose birthyears range from 1949 to 1989'. Also Alter ist konstant, Geburtsjahr variiert. Wenn Lautwandel vorkommt, dann müssten sich diejenigen mit hohem von niedrigem Geburtsjahr unterscheiden.

ii. 41 Sprecher geb. '1940 and 1949, who range in age from 30 to 67'. Dies ist noch eine Weise um Lautwandel festzustellen. Hier ist aber Geburtsjahr konstant, und Alter variiert.

iii. 18 Sprecher geboren zwischen 1890-1990 - also hier unterscheidet sich Geburtsjahr drastisch.

Die analysierten Wörter (S. 23)

'The current study focuses on monosyllabic, monomorphemic content words containing exactly one nasal segment extracted from the interviews with these 105 speakers. Word types are coded for position of the nasal segment relative to the vowel: preceding nasal, i.e., NV (e.g., mad), and following nasal, i.e., VN (e.g., home).'

Akustisches Parameter ist A1-A0. (Das hatten wir schon im Thema 9 oben, wie folgt. A1 ist die Amplitude vom F1 im Vokal, P0 ist die Amplitude vom nasalen Formanten. Je nasalisierter, umso kleiner A1 und umso größer P0. Daher wird das Verhältnis A1/P0 kleiner, umso nasalisierter. Oder auch: A1 dB - P0 dB wird umso kleiner je nasalisierter der Vokal ist. Das ist so, weil A1 dB - P0 dB eine logarithmische Skala hat und  $\log(A1/P0) = \log(A1) - \log(P0)$ . Nicht näher auf diesen Teil eingehen.

Dann kann alles übersprungen werden bis S. 29

'The apparent-time model testing the effects of all possible influencing lexical factors ... suggests that, all things being equal, higher frequency words have greater degree of vowel nasality'. Und dazu Abb. 6 auf S. 30: A1-P0 nimmt ab (weniger koartikulatorische Nasalisierung im Vokal) je seltener das Wort ist (seltene Wörter = haben höhere Zahlen auf der x-Achse).

Dann wieder alles überspringen bis S. 31:

'The fourth discussion point from this study is that lexical frequency is a language-internal factor that influences degree of coarticulatory nasality in naturalistic speech. The significant main effect...with more frequent words showing greater nasal coarticulation, is notably present while a post-hoc correlation between lexical frequency and duration was not demonstrated.'

Die Bedeutung davon ist wie folgt:

1. Häufige Wörter werden hypoartikuliert.
2. Hypoartikulation in häufigen Wörtern lenisiert den Nasal in  $\tilde{V}N$  ('we speculate that word-final consonants undergo increased lenition as word frequency increases'), d.h. der alveolare Kontakt wird schwächer. Die  $\tilde{V}$  Koartikulation in häufigen Wörtern dagegen wird stärker ('more frequent words showing greater nasal coarticulation'). Sie wird stärker, eventuell weil in häufigen Wörtern, in denen N lenisiert wird, die Velumsenkungsgeste im Vokal früher vorkommt (word final consonants undergo lenition but velum gesture starts earlier).
3. Daher in häufigen Wörtern: (a) eine größere  $\tilde{V}$  Koartikulation gekoppelt mit (b) einem lenisierten N.

4. Wenn N stark lenisiert wird oder sogar getilgt wird, dann kann ein Hörer die  $\tilde{V}$  Koartikulation nicht mehr dem finalen N perceptiv zuordnen d.h. die wahrgenommene Nasalisierung bleibt am Vokal hängen (siehe Thema 9, 'sound change and the covariation of speech').
5. Das ist genau die Bedingung, die zur Phonologisierung der Nasalisierung führen kann (Latein manus -> French /mɛ̃/, main, Hand).
6. Das stellt daher - entgegen der Ansicht Junggrammatiker übrigens - eine Verbindung zwischen lexikalischer Häufigkeit und Phonologisierung im Lautwandel her.

Der Rest vom Paper kann weggelassen werden.

### # 13 Imitation and change

#### **babel14.labphon.pdf**

Es geht hier hauptsächlich darum, inwiefern die phonetische Imitation durch soziale Faktoren beeinflusst wird, und inwiefern dadurch phonetische Imitation mit Lautwandel verbunden werden kann.

S. 124: viele Studien zeigen, dass die Aussprache in der Kindheit sehr stark von der Gemeinschaft beeinflusst wird. Z.B. lernen Kinder mit bayerischen Eltern, die auf die Schule in Hamburg gehen, eher eine Hamburger-Aussprache usw. (die Gemeinschaft hat eine stärkere Wirkung auf die Aussprache als die Familie). Der Grund, weshalb das so ist, ist nicht ganz klar. Die Autoren deuten darauf hin, dass Kinder die Wahl haben, sich die Ausspracheform zu wählen ('children's ability to selectively model their speech patterns on preferred input from their peer groups...the linguistic representations accessed in speech production are not modelled on the aggregate of experiences received from perceptual input, but rather a selective slice of those experiences').

Erwachsene werden auch von neuen Aussprachevarianten beeinflusst, aber (S. 124/125) hier wählen Erwachsene nicht wie Kinder aktiv eine bevorzugte Aussprachevariante, sondern Imitation geschieht eher passiv durch Kontakt. Meine Studien zur Königin, die hier zitiert werden, sprechen eher dafür, dass die Anpassung an eine neue Aussprache passiv geschieht (weil die Änderungen in der Königen von Jahr zu Jahr so klein sind, dass sie kaum bemerkbar sind). Auf der anderen Seite zeigt Labov (1963), dass junge inselbewohnende Männer, deren traditionelle Arbeit durch die Zuwanderung von 'Mainlanders' bedroht wurde, sich dieser Mainlander-Aussprache weniger anpassten (aber das könnte auch sein, weil sie zu den Mainlanders wenig Kontakt hatten).

S. 125

Communication accommodation theory (CAT): Konvergenz/Imitation entsteht als Folge der *sozialen* Entfernung zwischen Sprechern. Hier eine kurze Zusammenfassung von Trudgill (2008) [auch in der Webseite vorhanden: [trudgill08.langinsoc.pdf](http://trudgill08.langinsoc.pdf)]

Trudgill befasst sich mit der Entwicklung von neuen Aussprachen - z.B. Australisches Englisch, Amerikanisches Englisch, usw. . Trudgill widerlegt die Idee, dass eine neue Aussprache aufgrund nationaler Identität entsteht (ist also gegen die Idee, dass das Bedürfnis, sich als Australier zu zeigen, zur Entwicklung von australisch-Englisch

beisteuerte), sondern, dass die neue Aussprache aus einer Mischung der Aussprachen der Zuwanderer entsteht:

'But why exactly should dialect contact result in dialect mixture?

It is purely a matter of who interacts most often with whom – a matter of density of communication... accommodation is not only a subconscious but also a deeply automatic process...linguistic accommodation is not driven by social factors such as identity at all but is an automatic consequence of interaction. it is this innate tendency to behavioral coordination, not identity, that is the very powerful drive that makes dialect mixture an almost inevitable consequence of dialect contact'

Also ist für Trudgill Dialekt-Mischung eine automatische Folge der gegenseitigen Imitation. Daher: Kinder von norddeutschen Eltern, die eine bayerische Schule besuchen, entwickeln sehr schnell eine süddeutsche Aussprache, weil sie tagsüber so viel Kontakt zu anderen Kindern mit einer süddeutschen Aussprache haben.

S. 126: Experimentelle Methoden in der Imitation - hier ein paar mehr Details dazu:

Typischerweise *Baseline, Auditory Input, Production, Comparison*

*Baseline*: eine Sprechergruppe (i) wird aufgenommen

*Auditory Input*: dieselbe Sprechergruppe hört auditive Sprechstimuli eines anderen Sprechers (j) der nicht Bestandteil der Gruppe (i) war. Es kann sein, dass die Sprache von (j) synthetisch manipuliert wird (zB durch Längung von VOT).

*Production*: dieselbe Sprechergruppe (i) wird wieder aufgenommen, nachdem sie die Stimuli von (j) wahrgenommen haben.

*Comparison*: meistens AXB. A ist ein Stimulus aus der Baseline, B ein Stimulus aus Production, X ein Stimulus aus dem Auditory Input. Eine andere Hörergruppe muss entscheiden, ob es größere Ähnlichkeiten zwischen AX oder BX gibt. Der zweite Fall (=mehr Ähnlichkeiten in BX) deutet auf Imitation hin, also, dass Production von dem Auditory Input beeinflusst wurde.

Manchmal wird Anpassung an X durch den Unterschied zwischen A und B gemessen: je größer der Unterschied in die Richtung von X, umso mehr wurden die Versuchsperson vom Auditory Input beeinflusst.

### Goldinger's Exemplar/Episodic Theorie

-----  
(Siehe bitte auch Thema #11 zur Exemplartheorie). Ein Wort im Lexikon setzt sich zusammen aus allen auditiven Eigenschaften des wahrgenommenen Wortes inkl. phonetischen Details und Sprecherinformationen. Jedes Mal, wenn ein Wort wahrgenommen wird, wird eine auditive 'Spur' (trace) mit solchen Information dem Lexikon hinzugefügt. Imitation kommt zustande, weil

(a) die Spuren der vor kurzem wahrgenommenen Wörter besonders salient sind und  
(b) weil das gleiche Lexikon für die Perzeption und Produktion der Sprache eingesetzt wird.

Imitation ist stärker je kleiner der zeitliche Abstand zwischen Wahrnehmung und Produktion, da die Salienz der vor kurzem wahrgenommenen Wörter mit der Zeit abnimmt.

Imitation von Wörtern die selten vorkommen (low statistical frequency) ist wahrscheinlicher, weil es für solche Wörter nicht so viele im Lexikon gespeicherte auditive Spuren gibt - insofern wird der Einfluss einer neuen auditiven Spur größer sein. Babel schlägt vor, es wird ähnlich für neue Stimmen sein: "What this model similarly suggests is that atypical voices are predicted to be imitated more, as listeners have less experience with such cognitively novel input and should show activation patterns similar to low-familiarity voices."

Babel (S. 127) vertritt die Theorie, dass die soziale Anpassung - also grob inwiefern sich zwei Sprecher mögen - eine Auswirkung auf die Imitation hat. S. 127: 'The more attractive female participants rated the model, they more they accommodated his vowels.'

Daher eine Erweiterung von Trudgills Modell: die neue Entwicklung von einem Dialekt und daher Lautwandel entstehen nicht nur wegen automatischer Imitation sondern werden auch von sozialen Faktoren gesteuert. [Obwohl ich nicht dieser Meinung bin, Trudgill auch nicht!].

#### Die empirische Analyse

-----

es soll getestet werden, inwiefern atypische und/oder attraktive Stimmen eher imitiert werden (atypisch im Bezug zu dem Befund von Goldinger, dass atypische Wörter eher imitiert werden). Die Methode war wie folgt:

#### Modellsprecher auswählen

.....

60 Sprecher (30 M, 30 F) produzierten die isolierten Wörter mit /i, u, a/ Vokalen auf S. 128.

30 Hörern (15 M, 15 F) wurde 15 dieser Wörter präsentiert. Sie mussten nach der 15en Präsentation ein Urteil geben, wie attraktiv die Stimme war.

Unabhängig davon wurde ihnen ein Wort pro Sprecher präsentiert. Es wurde pro Hörer gemessen, wie lang gebraucht wurde, um das Geschlecht des Modellsprechers zu erkennen. Kürzere Zeiten deuten eher auf gender typicality hin.

Vielleicht dann kurz die Ergebnisse, ohne ins Detail einzugehen, anhand von Fig.1 S. 131 besprechen.

#### Auditory naming = Imitationsexperiment

.....

*Baseline* (S. 132) 20 Sprecher (10 M, 10 F) produzierten dieselben isolierten Wörter (A) wie in Tabelle 1.

*Auditory Input*: Diese 20 Sprecher hörten die Wörter der 8 ausgewählten Modellsprecher (X)

*Production*: Shadowing - das Wort des Modellsprechers wurde nachgesprochen, sobald es wahrgenommen wurde (B)

*Comparison*: 150 Versuchspersonen nahmen an einem AXB-Experiment teil. Ist BX ähnlicher als AX?

## Ergebnisse (Analysis and results)

.....  
Ein ziemliches Durcheinander. Daher eher überfliegen.

Fig. 2. Hörer beurteilen BX ähnlicher als AX wenn die Balken in den Abbildungen über die 50% (gestrichelten) Linien sind. Wenn das so ist, kann davon ausgegangen werden, dass imitiert wurde d.h. der Sprecher änderte die Sprache in die Richtung des ausgewählten Modellsprechers.

Daher (Fig. 2):

- Es wurde imitiert (da alle Balken > 50% sind).
  - Frauen (Sprecherinnen) imitierten mehr als Männer (Sprecher). Daher ist rot > blau.
- Und
- Dieser Unterschied zwischen Frauen und Männer war etwas ausgeprägter, wenn die Stimme des Modellsprechers weiblich war (daher Interaktion: der Unterschied zwischen rot und blau ist größer links als rechts).

Fig. 3.

- Typische Sprecher wurden imitiert (die Balken rechts sind sig. > 50%).
- Attraktivere Sprecher wurden von Frauen, aber nicht von Männern imitiert.

Fig. 4

Wie auf S. 134 erklärt, gab es zwei Gruppen von X-Modellsprechern.

(a) die more-convergence-Gruppe: das sind die attraktivsten und untypischsten Modellsprecher.

(b) die less-convergence-Gruppe: die unattraktivsten und typischsten Sprecher.

Wie der Name schon sagt, wurde mehr Imitation zu (a) als (b) erwartet.

Die Ergebnisse. Das war nur so für männliche Sprecher: der blaue Balken links > der blaue Balken rechts. Und die Ergebnisse zeigen wieder, was wir schon gesehen haben: Frauen imitierten mehr als Männer (daher rot > blau)

Abb 5 bis Ende der Ergebnisse würde ich weglassen - es ist wieder ein ziemliches Durcheinander.

Dann zur Diskussion um die Ergebnisse zusammenzufassen.

- i. Frauen imitierten mehr als Männer (wie wir oben gesehen haben)
- ii. Die untypischsten Stimmen wurden häufig imitiert. Jedoch imitierten Frauen auch die typischsten Stimmen.
- iii. Frauen imitierten die attraktivste Stimme mehr als die unattraktivste.

Babel sagt: In general, our evidence supports both the social preference and cognitive reflex hypotheses (S. 144) = die Stärke der Imitation richtet sich danach, wie attraktiv die Stimme ist, und wie untypisch sie ist. Meiner Meinung nach ist diese Interpretation gewagt. Ihre Ergebnisse zeigen nur dass (a) in einem solchen Shadowing-Experiment meistens imitiert wird und (b) Frauen das mehr tun als Männer - vielleicht haben Frauen die Modell-Sprecher näher/besser/aufmerksamer zugehört als die Männer es gemacht haben. In der Tat sagt sich auch so etwas S. 145 'Females may pay more

attention to the details of the input'. Ich finde ihre Argumente bez. societal and local power relations, and gender roles möglich, aber die Interpretation geht zu weit - aufgrund der vorhandenen Ergebnisse.

Der experimentelle Aufbau des Papers ist ganz gut, aber die Ergebnisse geben nicht so viel her: sie zeigen nur, dass in einem solchen Experiment imitiert wird und das Frauen das mehr tun als Männer.

### **babel14.pdf**

Das relevante aus diesem Paper beginnt S. 315, Spalte 2:

'Speech Production

The focus of this section is on studies of phonetic accommodation

Es gibt einen Überblick der wichtigsten Studien zur Imitation.

Insbesondere.

(a) Soziale Attraktivität - siehe auch (f) unten

.....

i. Die Studien von Natale (1975)

<http://pms.sagepub.com/content/40/3/827.full.pdf>

Paare (dyads) nahmen an einer Unterhaltung teil. Die Gesprächspartner waren einander nicht bekannt. Während der Unterhaltung wurde die akustische Konvergenz gemessen: hauptsächlich anhand der Dauer der Gesprächspausen beim Gesprächspartnerwechsel (inwiefern die Gesprächspartner dazu neigten, ähnliche Pausendauern zu haben).

Zusätzlich wurde 'social desirability gemessen' - also soziale Attraktivität.

Eine Korrelation konnte zwischen Konvergenz und soziale Attraktivität festgestellt werden.

ii. weitere Studien von Guy (S 316). Die Konvergenz in Telefongesprächen wurde gemessen u.a. anhand von Anpassungen im Langzeitspektrum. Zusätzlich haben andere diese Gespräche für 'smoothness' beurteilt, das bedeutet ca. wie kooperativ und flüssig ihnen das Gespräch vorkam. Eine Korrelation zwischen akustischer Konvergenz und 'smoothness' wurde gefunden.

(b) inwiefern der Sprecher seine Sprache dem wahrgenommenen höheren Status des Gesprächspartners anpasst.

.....

'Gregory and Webster's analysis indeed found that King accommodated more toward high status guests than lower status guests.'

(c) Mehr Konvergenz in tief- als hochpass gefilterte Sprache

.....

weil in tief-pass gefilterter Sprache die niedrigen Frequenzen und dadurch sehr viele prosodische Informationen, die für die Übertragung von Emotionen wichtig ist, erhalten bleibt. Und in hoch-pass gefilterter Sprache war der Blickkontakt und visuelle

Interaktion (Gestik usw.) beeinträchtigt (S. 23) - wahrscheinlich weil Blickkontakt für die Interaktion zwischen 2 Gesprächspartnern wichtig ist, und weil aber die Interaktion gebremst wird, wenn die Information in tiefen Frequenzen nicht vorhanden ist.

(d) Die Konvergenz hängt eventuell vom Geschlecht ab

.....  
 - konsistent mit dem Ergebnis in pardo14.labphon.pdf. In Pardo (2006) wurde die Sprache von Gesprächspartnern in einem Map-task analysiert (das ist wenn die beiden eine nicht ganz identische Karte bekommen, und müssen gemeinsam zwischen zwei Zeichen auf der Karte einen Weg finden; einer leitet (gives instructions), der andere folgt (receives instructions). Das Ergebnis: 'Female dyads were found to converge toward the speaker who was receiving instructions, whereas male dyads patterned oppositely; they converged toward the speech of the male talkers giving instructions.' (Vielleicht dann die anderen Pardo-Studien überfliegen/weglassen).

(e) Goldinger (1998). seltene Wörter werden imitiert (siehe oben).

.....  
 Siehe auch pdf S. 2, Spalte 2 vom nächsten Artikel nielsen14.jshlr.pdf 'Exemplar-based theories posit that every stimulus leaves a unique memory trace...'

(f) Die Konvergenz wird eher in einigen als in anderen Sprachlauten gezeigt.

.....  
 z.B. eine verlängerte VOT in stimmlosen Plosiven wird oft imitiert, und auch generalisiert. zB. Nielsen (2011) führte auch ein AXB Experiment durch, indem die VOT in stimmlosen /p/-Lauten gelängt wurde. Sprecher produzierten dann längere VOT-Werte nicht nur in /p/, sondern auch /k/ Lauten (=Generalisierung auf eine andere Artikulationsstelle), nachdem sie diese gelängte VOT im /p/ gehört hatten. [Aber eventuell ist es nicht, dass Sprecher eher längere VOTs imitieren, sondern dass es schwieriger ist, akustische Massnahmen bei anderen Lauten einzusetzen, um Imitation überhaupt festzustellen]. Siehe aber auch Absatz 3 S. pdf Seite 2 von dem nächsten Paper nielsen14.jshlr.pdf 'A growing body of research has revealed that phonetic imitation can be manifested in various phonetic features...'

(g) Eventuell mehr Konvergenz wenn sich die Gesprächspartner mögen.

.....  
 (Greenwald, McGhee, & Schwartz, 1998). Neuseeländische Sprecher mit einer positiven Einstellung zu Australien imitierten eher australisch-englisch als es Sprecher mit einer neutralen oder negativen Einstellung gemacht haben.

(Die Studie von Babel 2012 würde ich eventuell nicht erwähnen; die Ergebnisse sind zu komplex und vage um in 2 Zeilen zusammengefasst zu werden; aber das Paper ist babel12.jop.pdf in der Webseite).

Sehr wichtig und dies ist auch meine Meinung:

S. 317 ganz oben: 'We can gather from this research that imitation and convergent behavior appear to be, perhaps, an inevitable phenomenon... imitation is a pervasive behavior across human behavior in general'

und S 318 'Indeed, recent modeling work demonstrates that convergence is inevitable' Woran ich sehr zweifle ist: 'the fact that social factors mediate low-level behavior, such as imitation, suggests that speech production is never without social influence'. Erstens ist es nicht nicht eine \*Tatsache\* dass 'social factors mediate low-level behavior, such as imitation' - es gibt nicht genug Evidenzen dafür. Und zweitens ist sicherlich richtig 'speech production is never without social influence' aber das hat eventuell wenig mit Imitation zu tun!

Zum Schluss weisen (S. 319 oben) die Autoren mit Recht darauf hin, dass der Grad der Imitation zwischen Sprechern sehr unterschiedlich sein kann.

Das könnte etwas mit der auditiven Rückkoppelung zu tun haben, d.h. inwiefern man in der gesprochenen Sprache von der Wahrnehmung der eigenen Stimme beeinflusst wird. In solchen Experimenten hört ein Sprecher seine eigene Stimme via Kopfhörer. Mit einem Gerät dazwischen können die Formanten in Echtzeit verschoben werden, sodass man produziert z.B. 'Dank' hört aber 'Denk' weil F1 nach unten geschoben wurde (wir haben so ein Gerät im Institut und führen ein DFG-Projekt damit durch). Manche Sprecher kompensieren für diese akustische Verschiebung (sie öffnen den Mund weiter damit sie trotzdem 'Dank' hören) andere eben nicht. Eventuell könnte diese Variabilität in dem Einfluss der auditiven Rückkoppelung mit der Variabilität in der Imitation verbunden sein, obwohl das noch nicht nachgewiesen wurde, soviel ich weiß.

nielsen14.jshlr.pdf

=====

Dies ist wichtig aus dieser Studie (Abstract):

'Furthermore, an age effect of phonetic imitation was observed; namely, children showed greater imitation than adults, whereas the degree of imitation was comparable between preschoolers and 3rd graders.' Also Kinder imitieren mehr als Erwachsene. Das könnte sein, weil laut Exemplartheorie sie weniger gespeicherte Exemplare haben, sodass ihre Aussprache sehr stark durch neue Stimmen beeinflusst wird. Das könnte auch erklären (a) wieso Kinder von dem Dialekt beeinflusst werden, in dem sie aufwachsen (die Kinder aus Hamburg, die auf die Schule in München gehen, hören sich wie süddeutsche an, auch wenn die Eltern eine norddeutsche Aussprache haben) und (b) wieso Sprachwandel sehr oft mit Ausspracheänderungen in Kindern verbunden ist.

Zusätzlich zur Exemplartheorie werden auf pdf S. 2 zwei weitere Theorien zur Imitation erwähnt - diese könnten weggelassen, oder sonst schnell überfliegen.

(a) Carol Fowler und gesture theory of speech perception. Dies ist eigentlich die artikulatorische Phonologie. Der Kern hier: artikulatorische Gesten werden direkt im Sprachsignal perzipiert. Daher wird imitiert (weil man im Signal die geänderten Gesten hört). Zusätzlich kann das Zeitintervall zwischen Wahrnehmung des Modellhörers und dessen Imitation so klein sein (im Shadowing Experiment), dass diese Imitation teilweise automatisch sein muss - was wieder eine Folge ist, dass (für die artikulatorische Phonologie) Perzeption nichts mehr als die direkte Wahrnehmung der Gesten im Signal ist (d.h. das Zeitintervall kann laut dieser Theorie nur deswegen so klein sein, weil Gesten direkt perzipiert werden).

(b) Communicative accommodation theory - hatten wir schon früher. Das ist mehr oder weniger, dass man sich die eigene Sprache an die Sprache des Gesprächspartners anpasst, um die soziale Distanz zum Gesprächspartner zu reduzieren.

Phonetic Imitation (pdf S. 3) ganz unten

dieser erste Absatz ist verwirrend 'Phonetic imitation is unlike other forms of imitation, in that phonetic features in speech signals are known to be perceived categorically.' Das ist gar nicht der Fall; oder eher die Aussage ist zu schwammig, sodass man nicht weiß, was sie damit meint. Daher diesen ersten Absatz eventuell vergessen.

Diesen Absatz 'Previous research in phonological acquisition has revealed that phonological categories...' usw eventuell auch weglassen. Das sagt nicht mehr als (a) die Phonologie (also Wahrnehmung von Kategorien) sich mit zunehmendem Alter stabilisiert und (b) die geringere Fähigkeit von Kleinkindern Kategorien wahrzunehmen, könnte vielleicht mit Imitation zusammenhängen.

Vielleicht dann zur Methode springen.

Kinder aus dem Kindergarten (5 Jahre), Kinder im Alter von ca. 8 Jahren, und Studenten nahmen an einem Imitationsexperiment teil.

Es ist wieder ein AXB Experiment.

X = Eine Modellsprecherin, die Wörter mit initialem /p/ produzierte. VOT wurde in diesen Wörtern künstlich gelängt.

A und B bestehen aus Wörtern produziert von den obigen 3 Gruppen. Diese waren sowohl /p/-initiale und /k/-initiale Wörter. Die /p/-initiale Wörter enthielten 'exposure words' (wenn sie von der Modellsprecherin produziert wurden) und 'nonexposure words' (/p/-initiale Wörter die von ihr nicht produziert wurden). Alle /k/-initiale Wörter waren 'non-exposure'.

Die Ergebnisse

Einfach Fig. 1 präsentieren und kommentieren also:

- i. Alle Gruppen hatten einen längeren VOT in B als in A (sonst wären die Balken bei 0 ms) - d.h. in allen Gruppen wurde VOT gelängt = sie wurden von der Modellsprecherin beeinflusst.
- ii. Das war so am meisten für exposure /p/-Wörter vor allem bei 8-jährigen Kindern und Erwachsenen
- iii. Der Effekt der Längung war größer in /p/-initialen als in /k/-initialen Wörtern in 8-jährigen Kindern und Erwachsenen
- iv. Das wichtigste: Die Längung war am größten für Kinder (5 und 8 Jährige) als für Erwachsene

Fig. 2 bestätigt auch ganz deutlich, dass VOT gelängt wurde. Denn wenn keine Längung in B vorgekommen wäre, müssten alle Punkte auf der Linie liegen. Sie liegen aber deutlich darüber - das zeigt, eine längere VOT in B als in A. Die Abbildung zeigt auch (a) dass VOT sehr zwischen den Sprechern variiert (manche haben generell eine kurze und manche eine längere VOT; dies ist wahrscheinlich weil Sprecher mit unterschiedlichen Sprechgeschwindigkeiten reden, und Sprechgeschwindigkeit beeinflusst Dauer-Variablen wie VOT und (b) der Grad der Längung ist auch recht sprecherabhängig. Manche imitieren kaum (Punkt liegt nah an der Linie) manche recht viel (Punkt liegt weit oberhalb der Linie).

## Diskussion

Eine der wichtigsten Zeilen ist:

'we observed phonetic imitation across all age groups, which indicates that fine phonetic details in the speech signal are always perceived and retained to some degree, even in adults.' (pdf S. 7 Spalte 2)

- Keine geschlechtsspezifischen Unterschiede
- Es kann nicht sein, dass die Imitation sich auf Phoneme beschränkt: denn dann wäre VOT nur in /p/ und nicht in /k/ gelangt worden. Statt dessen werden phonetische Merkmale wie Aspiration imitiert.

Only exemplar-based theories (e.g., Goldinger, 1998) readily predict the age effect observed in the current study, because younger speakers should have fewer exemplars than adults, and thus the weight of each exemplar would be relatively greater.

jedoch (S. 8)

no difference between preschoolers and third graders was observed, although the exemplar view would predict greater imitation for preschoolers who are expected to have fewer exemplars.

Alternativ (pdf S. 8, Spalte 2) lassen sich diese Ergebnisse im Rahmen vom 'The critical period hypothesis' (Johnson & Newport, 1989) erklären. Diese Erklärung geht auf Lenneberg (1967) [Biological Foundations of Language] zurück. Die Theorie ist, dass das Gehirn bis zur Pubertät genügend Plastizität hat, um neue Sprachen zu erlernen, sodass Kinder fast automatisch eine neue Sprache lernen, wenn sie sich in der Sprachgemeinschaft länger aufhalten. zB. englische Kinder, die nach Deutschland umziehen, lernen ziemlich automatisch und schnell deutsch und werden mehr oder weniger bilingual, \*aber nur unter einem gewissen Alter\* (= critical age). Vielleicht waren die Kinder dieser Studie deswegen für Imitation anfälliger, weil sie noch unter dem critical age liegen.

Daher 'the natural next step is to investigate a more fine-grained age effect of phonetic imitation by examining speakers from more age groups, including pre- and postpuberty.'

## # 14. Acquisition of speech by bilinguals

flegespecom2003.pdf

+++++

Hier wird der von Flege entwickelte SLM (speech learning model) getestet. Das SLM befasst sich mit der kognitiven Darstellung der Laute von zwei Sprachen: L1 (die Muttersprache) und L2 (die zweit erlernte Sprache).

Laut SLM interagieren die L1-L2-Lautsysteme. Als Folge dieser Interaktion.

(a) vor allem wenn zwei Laute in den beiden Sprachen phonetisch sehr ähnlich sind, dann kann ein einziger Laut, der Mittelwert der beiden, entstehen. z.B. deutsch und englisch /i/ sind sehr ähnlich (lief, leaf) aber nicht identisch: insbesondere ist der deutsche /i/ etwas periphärer (gespannter). Für einen L1-englisch-Muttersprachler oder einen L1-deutscher-Muttersprachler, der die andere Sprache lernt, könnte sich ein Mittelwert dazwischen bilden, der dann in beiden Sprachen verwendet wird. Insofern weicht dann für einen solchen Sprecher der /i/ ab, sowohl von dem /i/ im L2 und (interessanterweise) auch in der Muttersprache. Diese Bildung des Mittelwertes wird Assimilation genannt. Diese Theorie hat also zur Folge, dass das Erlernen einer zweiten Sprache die Laute in L1 - also in der eigenen Muttersprache verändern (sodass der Sprecher, der eine zweite Sprache erwirbt, nicht mehr genau wie ein Muttersprachler klingt).

(b) Es kann aber auch sein, dass die L1- und L2-Laute (auch wenn sie relativ ähnlich sind) sich gegenseitig abstoßen. Das wäre z.B. der Fall wenn für das obige Beispiel das deutsche /i/ noch periphärer (gespannter) und der englische /i/ etwas zentralisierter als in der Muttersprache produziert werden. Dies ist Dissimilation. In diesem Fall wird es auch der Fall sein, dass das Erlernen einer Zweitsprache zur Folge hat, dass man nicht mehr genau muttersprachlich klingt.

Diese Theorie wird anhand von bilingualen Sprechern (L1-Italienisch, L2-Englisch) geprüft.

Die Bilingualen Sprecher teilen sich wie folgt auf:

AOA = age of arrival = in welchem Alter sie nach Kanada ankamen  
usage: low/high. low = wenig L1 (Italienisch) gesprochen; high = viel L1-Italienisch gesprochen.

Die wichtigsten 2 Gruppen hier:

Early-low. Im jungen Alter nach Kanada ausgewandert, und wenig L1 gesprochen. Diese Gruppe lernt/erkennt den Unterschied zwischen dem italienischen Monophthong und dem britischen /ei/ Diphthong. Da sie den Unterschied erkennen, und lernen, kommt es hier zur Dissimilation d.h /e:, ei/ stoßen sich gegenseitig ab. Dies erfolgt durch die Produktion eines sehr ausgeprägten /ei/ Diphthongs - d.h. der Diphthong ist für diese Gruppe ausgeprägter als für L1-englisch-Muttersprachler.

Late-high. In einem höheren Alter nach Kanada ausgewandert, und viel L1 gesprochen. Sie sind nicht so sehr imstande italienisch /e:/ von englischem /ei/ zu trennen (können nicht getrennte Kategorien bilden). Sie mitteln daher die Kategorien. Daher: viel weniger diphthongisiert als bei den L1-englisch-Muttersprachlern.

Die Methode/Ergebnisse in diesem Paper sind recht umständlich. Die Abbildungen sind auch ziemlich schlecht. Vielleicht das wichtigste hier:

Fig. 4: /ei/ ist deutlich mehr diphthongisiert in englisch als in italienisch.

Fig. 5 S. 484. Die Linie ist länger für early-low als für NE als für Late-high.

Das zeigt also, dass early-low mehr diphthongisiert.

Die größere Diphthongisierung kam nicht deswegen zustande, weil Early-Low langsamer sprach (S 484 unten)

Keineswegs auf die Details näher eingehen.

Die allgemeine Schlussfolgerung

S. 486. only participants in an Early-low group were able to discriminate English /ei/ and Italian /e/ tokens at a significantly above-chance rate. Participants in the two late bilingual groups, on the other hand, may have produced English /ei/ with less movement than the NE speakers did because many of them continued to treat /ei/ tokens as instances of the Italian /e/ category, and so merged the properties of English /ei/ and Italian /e/.

S. 487 (das gleiche): The first prediction was that participants in the Early-low group would produce English /ei/ with more movement than participants in the two late bilinguals groups (Late-low, Late-high) would. The second prediction was that the late bilinguals would produce English /ei/ with less movement than the NE speakers. Both hypotheses were supported... Participants in the two late bilingual groups may have tended to produce English /ei/ with less movement than the NE speakers did because they continued to treat /ei/ tokens as instances of the Italian /e/ category. That is, they may have merged the properties of English /ei/ and Italian /e/ as the result of the mechanism of category assimilation.

BurnsWerkerMcVie2003.pdf

+++++

Ein schönes Paper - typisch für Werker, sehr sauber, logisch, und klar strukturiert - mit wichtigen Befunden für den Spracherwerb.

Es geht hier um 'perceptual narrowing'. d.h. Kleinkinder (KK) unter 6 Monaten hören Unterschiede zwischen Lauten, die in der Muttersprache nicht vorkommen; dann ab ca. 12-14 Monaten hauptsächlich phonetische Unterschiede, die kontrastiv (phonologisch) sind (daher 'perceptual narrowing', weil diese Fähigkeit, den Unterschied zwischen allen Sprachlauten der Welt wahrzunehmen, verloren geht).

Das obige ist meistens nur für monolinguale jedoch nicht für bilinguale getestet werden. Daher diese Untersuchung.

Drei Modelle werden getestet

(a) von Flege (siehe oben). Bilinguale KK haben ein einziges System für beide Sprachen mit der Folge, dass sie sich von Monolingualen in beiden Sprachen abweichen.

(b) von Grosjean. Bilinguale KK bilden zwei getrennte Lautsysteme für L1 und L2. Sie sind wie monolinguale in beiden Sprachen.

(c) wie bosch2003LS.pdf unten - KK sind wie monolinguale in nur einer der beiden Sprachen.

### Die allgemeine Methode

+++++

Der Vergleich ist zwischen englischen Monolingualen und engl./franz. Bilingualen.

Es geht hier um ein Habituation (siehe unten) auf unaspiriertem [pa] und es wird getestet ob KK den [ba, pa] sowie den [pa, pha] Unterschied wahrnehmen. (pha = aspirierter /p/).

Hier ist wichtig zu bemerken: L1-englisch Erwachsene differenzieren [ba, pa] von [pha] (= hören den Unterschied nicht zwischen [ba, pa] und hören den Unterschied zwischen [ba, pha] und [pa, pha]); L1-franz. Erwachsene differenzieren [ba], von [pa, pha].

### Mehr zur Methode

.....

visual habituation.

Hier eine Zusammenfassung

<http://neuwritesd.org/2013/11/18/antidepressants-plasticity-and-language-development/>

Fig. 2

Das ist alles eine Erweiterung der Methode in der visuellen Wahrnehmung: <https://www.youtube.com/watch?v=dlilZh60qdA>

Text dazu: [http://artsites.uottawa.ca/zamuner/doc/Johnson\\_Zamuner\\_EMLAR.pdf](http://artsites.uottawa.ca/zamuner/doc/Johnson_Zamuner_EMLAR.pdf)  
(Visual fixation procedure)

- (1) Ein checkerboard wird beleuchtet und gleichzeitig wird [pa] mehrmals abgespielt.
- (2) Das KK schaut hin, und irgendwann mal weg. Sobald es wegschaut, schaltet sich der checkerboard aus, und [pa] wird nicht mehr abgespielt.
- (3) Die Vorgänge (1, 2) werden mehrmals wiederholt. Das Kind schaut für eine immer kürzere Zeit hin. Irgendwann stabilisiert sich diese Zeit des Hinguckens. (z.B. nach dem 10en oder 20en Mal schaut das KK immer für ca. 1,5 Sekunden hin). Das ist der Habituation time.
- (4) Hier wird (1) wiederholt aber mit einem anderen Laut: [ba] oder [pha] (aspirierter /p/). Wenn das KK einen Unterschied zwischen den Lauten in (1) und (4) hört, dann schaut es länger hin = z.B. für 4 Sekunden. Dishabituation time wäre dann  $4 - 1.5 = 2.5$  Sekunden.

Die Frage ist dann, ob dieser Dishabituation-Time signifikant größer ist als Null (= unterscheiden sich die Zeiten in (1) und (4) signifikant?). In Fig. 1 sieht man einen langen (signifikanten) dishabituation time für [b] (grüner Balken ist hoch) aber nur einen kleinen (nicht-signifikanten) für [pha] (ganz rechter Balken). Das ist so für beide Gruppen. d.h. KK hören den Unterschied zwischen [ba, pa], aber nicht zwischen [pa, pha]. Daher die Abbildung unten mit dem grünen Strich und 'all 6-8 month olds'.

### Die wichtigsten Ergebnisse

.....

Expt. 1 Keine Unterschiede zwischen monolingualen und bilingualen KK-6. Beide Gruppen kategorisieren [ba, pa] vs [pha]. Es gibt keine Unterschiede zwischen monolingualen und bilingualen Kindern, wahrscheinlich weil das phonologische System noch nicht reif genug ist (eigentlich müsste daher [pa, pha] trennbar sein, aber das war nicht der Fall, eventuell aufgrund Probleme mit der Synthese).

Expt. 2 Hier gibt es bei der monolingualen Gruppe eine Phonologisierung - d.h. eine Kategorisierung [ba, pa] vs [pha] wie es englische Erwachsene tun. Das ist jedoch nicht der Fall für die bilingualen KKs. Dies ist das erste Zeichen, dass die Entwicklung bei mono- und bilingualen KKs anders abläuft.

Expt. 3. 14-21 Monatige. Hier jetzt 3 Gruppen. (a) Monolinguale-Englisch, (b) Monolinguale-Franz., und (c) Bilinguale. Die Frage: ist (c) ähnlicher wie (a) oder ähnlicher wie (b) oder macht (c) ganz was anders?

Wie zu erwarten, kategorisieren (a) und (b) wie Erwachsene in den jeweiligen Sprachen es tun.

Die Ergebnisse bei bilingualen Kindern ist viel komplizierter. Um es kurz zu halten.

- i. einige kategorisieren wie L1-englische KKs: [ba, pa] vs [pha]
- ii. einige wie L1-franz. KKs: [ba], vs [pa, pha]
- iii. einige machen den 3 fachen Kontrast: [ba] vs [pa] vs [pha]

i, ii wären mit bosch2003LS.pdf konsistent  
iii ist was Grosjean vorhersagt.

bosch2003LS.pdf

+++++

Es geht hier wieder um perceptual narrowing und Bilingualismus, auf eine ähnliche Weise wie im Paper oben. Hier werden aber katalanisch und spanisch geprüft bez. eines Kontrastes, das im katalanischen jedoch nicht im spanischen vorkommt (der /e,ε/ Kontrast).

Es gibt 3 Gruppen:

S: spanische-monolinguale (S4 bedeutet: spanische Monolinguale im Alter von 4 Monaten usw.)

K: katalanische-monolinguale: K

SK: spanisch-katalanische-bilinguale: SK

Alle 3 Gruppen im Alter von 4 Monaten konnten auf eine ähnliche Weise den /e,ε/ Kontrast wahrnehmen, da es noch keine Phonologisierung gibt.

Nur K8 und nicht S8 konnten den /e,ε/ Kontrast wahrnehmen - wie zu erwarten, da der /e,ε/ Kontrast im Katalanischen jedoch nicht im Spanischen vorkommt. Wider Erwartungen konnte SK8 diesen Kontrast \*nicht\* wahrnehmen.  
Für bilinguale war der Kontrast erst ab 12 Monaten wahrnehmbar.

Laut Autoren.

Die Abnahme der /e,ε/ Wahrnehmung für SK8 kommt vielleicht vor, weil akustisch das katalanische /e,ε/ und spanische /e/ miteinander überlappen (das spanische /e/ ist scheinbar akustisch ca. zwischen dem katalanischen /e,ε/) und weil das spanische /e/ sehr häufig vorkommt.

Die Erklärung für die spätere Differenzierung des /e,ε/ Kontrastes im Alter von 12 Monaten: vielleicht (S. 238) in dem sich die verschiedenen Lexika in den beiden Sprachen entwickeln; oder einfacher (S. 239), weil bilinguale Kinder zunehmend mit dem /e,ε/ Kontrast Erfahrung gewinnen.

sancierjop1997.pdf

+++++

Die Hauptuntersuchung ist die Aussprache einer bilingualen Frau (Portugiesisch, Amerikanisch-Englisch) mit Wohnort in den Vereinigten Staaten, und ob sich ihre Aussprache nach einem Aufenthalt in Brasilien verändert. Die Analyse befasst sich mit VOT, das im brasilianisch-portugiesischen (BP) viel kürzer ist als im amerikanisch-englischen (AE). Die Frage ist dann, ob VOT im AE nach dem Aufenthalt in Brasilien kürzer wird (und umgekehrt).

Die Sprecherin produzierte

BP-Sätze aus dem englischen übersetzt [BP]

AE-Sätze aus dem amerikanisch-englischen übersetzt [AE]

d.h. ihre Sprachaufnahmen waren immer gleichzeitige Übersetzungen der gelesenen Sprache. [BP, AE] wurden in drei Sessions aufgenommen:

- (a) Session 1: nach einem 4.5 monatigen Aufenthalt in USA unmittelbar vor der Reise nach Brasilien
- (b) Session 2: gleich nach einem 2.5 monatigen Aufenthalt in Brasilien
- (c) Session 3: eine Wiederholung wie in Session 1

Expt 1

13 L1-BP-Hörer mussten beurteilen, welche der [BP] (portugiesisch produzierten) Sätze aus Sessions (b) oder (c) englisch-gefärbter waren.

Ergebnisse

Eher (c) also nachdem die Sprecherin sich 4.5 Monate in den USA aufgehalten hatte

Expt 2

das gleiche aber diesmal mit den Sprachen in der anderen Reihenfolge also 33 L1-AE-Hörer mussten beurteilen ob die [AE] (englisch gesprochene) Sätze aus Sessions (b) oder (c) portugiesisch gefärbt waren.

Ergebnisse

Es gab keinen Unterschied in der Beurteilung von (b) oder (c).

Daher scheinen BP-Hörer die englischen Einflüsse auf Portugiesisch nach dem Aufenthalt in den USA zu erkennen; aber die AE-Hörer erkennen nicht die portugiesischen Einflüsse auf Englisch nach dem Aufenthalt in Brasilien.

### Expt 3

Vielleicht kommen die unterschiedlichen Ergebnisse aus Expt 2 vor, weil nach den Aufenthalten phonetische Veränderungen im Portugiesischen, jedoch nicht im Englischen vorliegen. Dies wurde geprüft durch VOT-Messungen von /p, t/ in den beiden Sprachen.

Die Erwartung: nach einem Aufenthalt in Brasilien müsste VOT vom BP /p, t/ größer werden; und nach dem Aufenthalt in den vereinigten Staaten müsste VOT vom AE /ph, th/ kleiner werden.

Fig. 1. Dies war so für BP /p t/ (Spalte 1). Hier müssten die linken und rechten Balken höher sein als der weiße Balken in der Mitte (was der Fall ist).

Dies war zum großen Teil auch so für AE /ph, th/. Hier müsste der mittlere Balken kleiner sein als die linken und rechten in Spalte 2 in derselben Abbildung. Abgesehen vom Vergleich BP-Aufenthalt mit US-2 Aufenthalt für /th/ war das der Fall.

Die Ergebnisse werden durch Imitation erklärt, und auch dass die Laute der beiden Sprachsysteme aneinander im Sinne von Flege gekoppelt sind.

(Die letzten 2 Seiten ab "We suspect that deeper principles might be found in the dynamical systems perspective" und ihre nicht ganz überzeugende Erklärung im Rahmen von artikulatorischer Phonologie/Action theory kann weggelassen werden).

### # 15. Complexification and simplification in languages

Beide Kapitel sind meines Erachtens sehr gut und mit vielen interessanten und nützlichen Informationen. Bitte aber keineswegs alle Details darstellen, sondern nur die wichtigsten und ggf. nur kurz mit einigen Beispielen beleuchten.

trudgill2012a.pdf

+++++

Die Frage ist, ob eine soziolinguistische Typologie - d.h. die sozialen und demographischen Eigenschaften einer Sprechgemeinschaft - die Komplexifizierung und Vereinfachung in der Morphologie erklären können. Das wichtigste. Sowohl Komplexifizierung (K) als auch Vereinfachung (V) entstehen aufgrund von Kontakt zwischen Sprachen oder Dialekten. Aber (V) entsteht hauptsächlich in Kreolsprachen aufgrund von einem kurzfristigen Kontakt (in 1-2 Generationen zB) zwischen Erwachsenen; und (K) eher aufgrund von langfristigem Kontakt (über mehrere Jahrhunderte) zwischen Kindern, die auch Bilingual sind. (K) entsteht also wenn Gesellschaften von zwei oder mehreren Sprachgruppen allmählich zusammenwachsen.

Das obige kommt wegen eines externen Kontaktes mit einer anderen Gemeinschaft zustande. Aber (K) kann auch intern entstehen - indem sich eine kontaktarme, sozial eng verbundene (socially tightly knit) Sprache komplexifiziert - auch eventuell über mehrere Jahrhunderte.

Einige weitere Details dazu wie folgt.

S15

Sprachen unterscheiden sich in ihrer linguistischen Komplexität - die Morphologie ist ein gutes Indiz für Komplexifizierung.

Hockett: alle Sprachen haben eine ähnliche Komplexität. Aber Kreolsprachen sind doch laut Trudgill in vielen Hinsichten einfacher

S16

es ist, contra Hockett, nicht der Fall dass die Vereinfachung in einem Teil der Linguistik (zB Syntax) durch die Komplexifizierung in einem anderen (zB Morphologie) ausgeglichen wird.

Koiné ausgesprochen [koine] = der Dialekt, der aus einer Mischung von 2 oder mehreren Dialekten entsteht (wie in neuseeland-englisch).

Es gibt viele Evidenzen, dass Kontakt mit morphologischer Vereinfachung verbunden ist. zB wenn man die Morphologie vom isländischen 'reich' mit den entsprechenden Formen im Norwegischen vergleicht (S. 20).

S. 21

Wie wird morphologische Vereinfachung gekennzeichnet?

i. Regularisierung

ii größere morphologische und syntaktische Transparenz - zB selten -> nicht oft usw.

iii Redundanz fällt weg. Sprachen sind inhärent redundant (dadurch das zB in deutsch 2 pers. sing. sowohl mit 'du' also auch mit '-st' gekennzeichnet ist. Eventuell kommt das vor, um die Wahrscheinlichkeit zu erhöhen, dass die Bedeutung übertragen wird (ohne -st müsste man sich ausschließlich auf 'du' verlassen, um 2pers.sing. zu erkennen).

Wenn eine Sprache vereinfacht wird, geht diese Redundanz verloren. Es gibt zwei Sorten von Redundanz.

a. Syntagmatische Redundanz. zB dass Geschlecht sowohl im Artikel als auch im danach kommenden Adjektiv signalisiert wird zB 'einen guten Mann'

b. Paradigmatische Redundanz

Die morphologische Markierung von Kategorien im selben Wort vor allem Genus (der/die/das), Kasus (guter/guten/gute/gutem; Du/Dich); Anzahl (Dich/Euch) usw.

Wenn paradigmatisch vereinfacht wird, dann ersetzt zB eine Präposition oft Kasus wie im altenglischen, godum huntan -> to a good hunter usw. Das muss aber nicht so sein. zB ist im englischen die morphologische Markierung von Genus weggefallen, ohne dass es kompensiert wurde.

S26

synthetic/analytic structures.

Synthetic = Eine synthetische Sprache ist eine mit vielen morphologischen Flexionen; analytic = wenig. Daher ist deutsch synthetischer als englisch.

Komplexifizierung

.....

wenn Merkmale von einer Sprache zu einer anderen hinzugefügt werden (=additive complexity).

S. 28

dependent marking und head marking (marking bedeutet Beschriftung zB 2nd pers sing ist in gehst 'markiert' durch -st. (-st marks the 2nd person sing).

head = der obligatorische Teil (im syntaktischen Sinne) einer Phrase.

zB alle Verbphrasen ('geht', 'geht nach Hause') brauchen einen Verb. Daher ist der Verb der 'Head' der Phrase (weil der Verb obligatorisch ist).

Head-marking ist wenn der Head durch die Morphologie geändert wird.

zB ich gehe/du gehst/er geht = head-marking, wegen Flexionen im Head 'geh-'

Dependency-Marking ist wenn die anderen Konstituenten der syntaktischen Phrase morphologisch variieren. zB Diese Bücher ist ein Noun-Phrase, der Head = der Nomen = Bücher, es gibt dependency-marking in 'diese' (man braucht hier wegen der Mehrzahl 'diese' und nicht 'dieses').

Mehr Info:

[http://www.uni-regensburg.de/sprache-literatur-kultur/sprache-literatur-kultur/allgemeine-vergleichende-sprachwissenschaft/medien/pdfs/artjh\\_2001\\_headmarkingdependentmarking.pdf](http://www.uni-regensburg.de/sprache-literatur-kultur/sprache-literatur-kultur/allgemeine-vergleichende-sprachwissenschaft/medien/pdfs/artjh_2001_headmarkingdependentmarking.pdf)

Eine größere Komplexität kommt hauptsächlich wegen Sprachkontakt zustande. (Der Unterschied zwischen residual and spread zones vielleicht weglassen). Wichtiger" 'independently of whether they are in residual or spread zones... almost all these high-complexity languages are in areas of considerable linguistic diversity and contact. Siehe dann Beispiel S. 30 von Tariana/Tucano

S. 32-32. Es ist ein Paradox, weil Kontakt scheinbar gleichzeitig zu Vereinfachung und Komplexität führen kann.

Die vorgeschlagene Lösung (S. 34)

Vereinfachung: wenn der Kontakt wegen Erwachsenen entsteht.

Komplexifizierung: wenn der Kontakt aufgrund von Kindern entsteht.

d.h. S. 35 die Entwicklung einer Pidgin-Sprache - also einer Sprache, die keine Muttersprache ist, sondern für Kommunikation zwischen Leuten verschiedener Sprache verwendet wird - führt zu Vereinfachungen und wird meistens von Erwachsenen gelernt/zwischen Erwachsenen übertragen.

Es kommt aber zu keiner Vereinfachung, wenn Sprachkontakt durch Kinder vermittelt wird. Der Grund: weil Kinder unter einem gewissen Alter ohne große Mühe eine Sprache erlernen; sie lernen vor allem die Aussprache. Das geschieht nur laut dem 'critical age hypothesis', wenn Kinder unter einem gewissen Alter sind (meistens vor der Pubertät).

S. 38/39. L2-difficulty (Zweitspracherwerb)

Das Argument bislang:

die Interaktion zwischen Sprachen aufgrund von Erwachsenen hat meistens eine Vereinfachung zur Folge, weil Erwachsene (im Gegensatz zu Kindern) große Schwierigkeiten haben u.a. morphologische Komplexität zu lernen. d.h. Sprachen mit hoher morphologischer Komplexität sind schwierig für Erwachsene, die eine Fremdsprache lernen (S. 38).

Dies bedeutet dann auch, dass die Flexionsmorphologie als erstes verschwinden müsste, wenn Sprachkontakt aufgrund von Erwachsenen zustande kommt.

Agglutinative Sprache: wenn jedes Morphem eine eigene Bedeutung hat. zB Türkisch: adam-lar-dan: lar = Mehrzahl, dan = Ablativ.

Fusional language: wenn ein Morphem mehrere Bedeutungen gleichzeitig vermittelt. zB Italiensch andrai (du wirst gehen): 'rai' überträgt gleichzeitig: Du, Singular, Futur, Indikativ.

Vergleiche mit deutsch Du+wirst+gehen eher agglutinativ und jedes Morphem (Du, wirst) hat eine eigene Bedeutung.

Agglutinative Sprache lernen Erwachsene schneller als fusionale Sprachen. Agglutinative Sprachen sind auch meistens analytischer (siehe oben - oder weniger synthetisch = mit weniger Flektionen).

Weitere Beweise, dass synthetische Sprachen komplexer sind:

- i. Pidgin-Sprachen sind typischerweise analytische Sprachen
- ii. Im Spracherwerb (Alter 2-5 Jahren) lernen Kinder morphologische Komplexität später (und machen oft in der frühen Kindheit morphologische Fehler).

Komplexifizierung kommt vor allem in der Situation des Sprachbundes (S. 42) vor, d.h. wenn genetisch unverwandte Sprachen längere Zeit in einer geographischen Region co-existieren, miteinander Kontakt aufweisen, und wo sich vor allem unter Kindern vor ein Bilingualismus entwickelt.

Ab Seite 43 ist nur folgendes wichtig:

Nubi (S. 44) hat eine morphologische Vereinfachung und S. 45 'it is well known that Nubi arose as a result of massive language contact'.

Alles andere vergessen bis eventuell S. 46 maltesisch, eine relativ isolierte Sprache, die eine morphologische Komplexifizierung aufweist. Die Komplexifizierung entstand durch die Einflüsse von Sizilien und von Süd romanischen (im Portugal/Spanien) Sprachen.

(das zusätzliche Beispiel aus zypriotisch kann weggelassen werden).

Altenglisch

.....

es geht hier um den morphologischen Wandel von fusional/inflectional in alt-englisch (450-1150) zur Vereinfachung der Morphologie im Mittelenglischen (1150-1500). Es wird vermutet, dass dieser aufgrund von Kontakt zustande gekommen ist.

Late British = Brittonic = Common Brittonic = Brittonic Late British (das ist alles das selbe!).

Brittonic/Late British war eine alte keltische Sprache, die in England zur Zeit von Alt-englisch eventuell bis zum 10 Jahrhundert. gesprochen wurde.

Es gab auch Einflüsse auf Alt-englisch von Altnordisch.

Der Kontakt zwischen Altnordisch und Alt-englisch war eher langfristig (über mehrere Jahrhunderte) zwischen Kindern und kann laut Trudgill daher nicht zu der Vereinfachung von Alt-englisch geführt haben, (S52): 'long-term co-territorial co-habitation and intermarriage... mixed marriage...cultural fusion of Angle and Dane in the north...'

Es gab vor allem viele Lehnwörter aus den Altnordischen, die bis heute noch existieren: 'they', 'them', 'their' usw.

Dies deutet auf einen langfristigen Kontakt hin - und daher Komplexifizierung durch 'additive borrowing' (aufgrund von Bilingualismus in Kindern).

Daher kann die Vereinfachung in Alt-englisch nicht aufgrund von Kontakt mit Altnordisch zustande gekommen sein.

Die Hypothese ist, dass es ziemlich viel Kontakt zwischen Brittonic (Late British) und Alt-englisch gab. Und auch dass dieser Kontakt eher zwischen Erwachsenen war (S. 53

und auch S 54: 'she suggests that for Britons (= keltische Sprecher von Brittonic) acquiring Old English, the first step was that of ... adult acquisition.'

Daher Vereinfachung durch Kontakt zwischen Erwachsenen.

#### Demography & Transmission

+++++

aber warum wurde dann die Morphologie von deutsch nicht vereinfacht, da es ja auch Evidenzen für Kontakt eher mit Erwachsenen anderer Sprachen gibt?

Trudgill hat eine (sehr) lange Antwort dazu, die wirklich nicht ins Detail besprochen werden muss. Das wichtigste, S 60 ganz unten. 'Contact between German and other languages has many times led to simplified forms... but this had no consequences for the mainstream language - the proportion of non-native speakers was nothing like high enough'

trudgill12b.pdf

+++++

Die Frage hier ist, inwiefern die Vereinfachung und Komplexifizierung, die eher für die Morphologie gezeigt wurden, sich auch in der Phonologie durchsetzt.

#### Vereinfachungen

.....

Pidgin-Sprachen haben oft kleinere Phonem-Inventare.

Bislama ist zB eine Pidgin-Sprache mit nur 5 kontrastierenden Phonemen. Aber es könnte sein (S. 118), dass das kleinere Phonem-Inventar von Bislama nicht aufgrund der Erwachsenen-gesteuerten Vereinfachung zustande kommt, sondern eher weil die Phonologie aus austronesischen Sprachen (mit einem kleinen Phonem-Inventar) übernommen wurde.

Zusätzliche Evidenzen für die Vereinfachung der Phonologie durch Kontakt: Wenn Dialekte aufeinander stoßen, dann kommt es oft zu Mergers (wie Rat/Rad = /ra:t/) - und Mergers führen notwendigerweise zu einer Verkleinerung des Phonem-Inventars. (Siehe auch Beispiel S. 118 in der Tabelle des Norwegischen).

#### Komplexifizierung

.....

Sprachen in einem Sprachbund nehmen gegenseitig Phoneme auf (additive borrowing), was zu einer Komplexifizierung des Phonem-Inventars führen kann - vor allem wenn das aufgrund von Bilingualismus in Kindern entsteht.

Ein zusätzliches Zeichen der Komplexität (S. 120-121) ist, wenn Laute phonologisiert werden (= für lexikalische Kontraste verwendet), die in den Sprachen der Welt recht selten sind (wie zB Klicklaute und vordere gerundete Vokale).

Bez. vorderer gerundeten Vokale argumentiert Trudgill, dass diese aufgrund von langfristigem Kontakt zwischen Kindern in genetisch unverwandten Sprachen entstanden sind. zB sind Breton und Französisch typologisch unverwandt (keltisch,

romanisch) und Breton hat eventuell Vokale wie [ø] aufgrund von Kontakt mit Französisch aufgenommen.

Die zusätzliche interessante Frage (S122) ist, ob Komplexifizierung \*ohne Kontakt\* entstehen kann. Dies soll laut Trudgill oft in kleinen, isolierten, 'tightly-knit' (Sprecher haben einen engen sozialen Kontakt zueinander) Sprachgemeinschaften vorkommen.

Es ist aber keineswegs der Fall, dass alle isolierten Sprachen - wie die polynesischen zB - komplexe Phoneminventare aufweisen. Hier ist Trudgill eventuell nicht ganz so überzeugend.

Er sagt, dass Sprachen wie Hawaiisch nur wenig Phoneme und daher wenig Silben hat. Aus dem Grund werden Wörter oft durch Reduplikation gebaut (S125, die Wiederholung ähnlicher Silben).

Und es ist sehr schwer für jemand, der diese Sprache als Fremdsprache lernt, diese phonologisch recht ähnlichen Silben zu differenzieren. (d.h. die Neighbourhood density ist sehr hoch in solchen Sprachen. Neighbourhood density = die Anzahl der Wörter, die sich von einem anderen Wort in nur einem Phonem unterscheiden).

Ab S. 126 'as far as the changes which led to...' bis S. 129 'Spontaneous complexification' sind die Argumente keineswegs überzeugend (bez. redundancy und der vermutlichen Verbindung zu einem kleinem Phoneminventar) und diese vielleicht daher weglassen. Es gibt aber einen wichtigen Punkt in diesen Seiten: dass Vereinfachung durch Kontakt zwischen Erwachsenen ziemlich schnell passiert; dagegen benötigt Komplexifizierung oft mehrere Jahrhunderte.

S. 129-132 Spontaneous complexification: sound change

----

der einzige wichtige Punkt hier: Lautwandel, die selten sind - d.h. die phonetisch schwer zu erklären sind - kommen oft in isolierten Sprachen vor.

Spontaneous complexification: vowel system

-----

Für lange Zeit wurde behauptet (S133), dass Vokale in den Sprachen der Welt symmetrisch verteilt sind (Crothers, 1978). Symmetrisch: Sprachen haben maximal distinktiv Vokale also /i, u, a/. Die neuere Forschung zeigt jedoch, dass viele Vokalsysteme asymmetrisch sind, und wenn das so ist, dann fehlt oft /u/. Dies wird auch von Schwartz et al (1997) gezeigt (S. 134): d.h. es gibt doch einige Sprachen, die eine asymmetrische Vokalverteilung haben. (Wir haben das auch mit /u/ belegt: <http://www.phonetik.uni-muenchen.de/~jmh/papers/jphon11.pdf>)

S. 134/135. Die Studie von Liljencrants & Lindblom (1972) hat durch ein computationelles Modell gezeigt, dass Vokale in der Regel in den Sprachen der Welt auditiv maximal differenziert sind: mit /i, u, a/ (= den drei maximal distinktiven Vokalen) dann /i, e, a, o, u/ = den 5 maximal distinktiven Vokalen usw.

Neuere Untersuchungen von australischen Sprachen zeigen jedoch, dass die Vokale nicht maximal sondern \*ausreichend\* verteilt sind. d.h. australische Sprachen mit 3 Vokalen

haben nicht /i, u, a/, sondern oft eher /ɪ, ʊ, ə/ (S. 135). Und dann gibt es (S135 unten) Sprachen mit asymmetrischen Vokalsystemen (in denen z.B. /e/ fehlt).

S136. Trudgill schlägt vor: eventuell ist die Theorie von 'maximal dispersion' entstanden, weil man im wesentlichen europäische Sprachen untersucht hat (in denen wie französisch oder italienisch die Vokale in der Tat eine maximale Verteilung haben). Aber vielleicht ist 'maximal dispersion' eher die Seltenheit. Wir wissen ungenügend dazu (S. 136), weil ungenügend zu isolierten Sprachen untersucht worden ist.

Trudgills weitere Hypothese: Maximal dispersion kommt eher vor in Sprachen mit hohem Kontakt; 'sufficient dispersion' eher in Sprachen mit niedrigem Kontakt. Die Begründung: In isolierten Sprachen mit einem geringfügigen externen Kontakt ist es wahrscheinlicher, dass Mitglieder der Sprechgemeinschaft sich gut kennen, und dass das Wortschatz und eventuell auch die Themen viel eingeschränkter sind. Dies hat zur Folge, dass viel mehr top-down processing (Kenntnisse der Sprache) eingesetzt werden kann, um eine Äußerung zu verstehen - d.h. in solchen Sprachen darf die gesprochene Sprache hypoartikulierter (im Sinne von Lindbloms H&H Theorie) als in Sprachen mit hohem Kontakt sein. In Sprachen mit hohem externen Kontakt muss jedoch die gesprochene Sprache eher verdeutlicht/hyperartikuliert werden, da Hörer meistens nicht so sehr imstande sind, top-down processing einzusetzen (die Themen und daher Wörter variieren mehr, und die Leute in einer Gemeinschaft mit sehr vielen externen Einflüssen sind einander nicht so gut bekannt - sodass man daher weniger hypoartikulieren kann, weil die Themen 'top-down' nicht so vorhersagbar sind).

Daher kommt sufficient dispersion eher in isolierten Sprachen mit wenig externem Kontakt vor.

Das ist ganz interessant und könnte auch stimmen. Aber eine andere Interpretation: wenn eine Sprache wenig Vokal-Phoneme hat, dann braucht man nicht den gesamten Vokalraum, um sie voneinander zu differenzieren: /ɪ, ʊ, ə/ reichen aus, wenn eine Sprache nur 3 Vokalphoneme hat. Der gesamte Vokalraum wird eher benötigt, wenn eine Sprache sehr viele Vokalphoneme hat (wie Deutsch).

Daher müsste untersucht werden, ob es der Fall ist, dass die Vokale in einer Sprache mit niedrigem Kontakt phonetisch aneinander näher sind im Vgl. zu einer Sprache mit hohem externen Kontakt, aber nur wenn die Anzahl der Phoneme in beiden Sprachen gleich ist.

Fast speech

+++++

S. 141-142. Die Hypothese hier ist, dass phonetische Vorgänge der schnell gesprochenen Sprache eher in isolierten, kleinen Gemeinschaften mit geringfügigem externen Kontakt vorkommen. Solche phonetischen Vorgänge der schnell gesprochenen Sprache sind vor allem oft dialekt-spezifisch (kommen nicht allgemein in Dialekten/Sprachen vor). d.h. Dialekte isolierter Sprachen haben oft idiosynkratische phonetische Merkmale - und das ist was das Erlernen eines solchen Dialektes für einen nicht-Angehörigen/Fremden der Sprachgemeinschaft so schwierig macht. Das ist laut Cohen (2012) <http://users.ox.ac.uk/~soca0093/pdfs/CohenCA2012.pdf> eventuell absichtlich so - denn

dadurch konnte man in einer kleineren Sprechgemeinschaft einen (eventuell gefährlichen) Fremden sehr schnell (an der Aussprache) erkennen können.

Phonologische Komplexität entsteht daher eher aus kleinen low-contact Sprachgemeinschaften, weil die Wahrscheinlichkeit größer ist, dass die idiosynkratischen phonetischen Vorgänge mit der Zeit phonologisiert werden. Um das klar zu stellen (denn dies ist sehr wichtig):

1. In kleineren Gemeinschaften mit wenig Kontakt wird oft hypoartikulierte, da die Mitglieder der Gemeinschaft die Bedeutung einer Äußerung sehr oft mit top-down Verarbeitung entschlüsseln können. (S. 142/143 'That is, the institutionalisation of fast speech phenomena into slow speech phenomena will be more typical of small tightly-knit communities where everybody knows everybody else and where there is a large fund of shared information - and where there are fewer formal situations').
2. Oft sind diese hypoartikulierten Formen dialekt-spezifisch und daher (eventuell absichtlich) schwer für einen Fremden zu imitieren.
3. Mit der Zeit werden diese hypoartikulierten Formen phonologisiert. Das ist also Lautwandel, in dem phonetische Vorgänge zum Zweck der Bedeutungsunterschiede eingesetzt werden.
4. Da aber die phonetischen Merkmale sehr idiosynkratisch sind, ist die Phonologisierung (und daher Erweiterung des Phonem-Inventars) oft in solchen kleinen, isolierten Sprachgemeinschaften sehr komplex. Komplex bedeutet hier eigenartig, d.h. Laute werden phonologisiert, die nur selten in anderen Sprachen vorkommen.

## # 16 The perceptual processing of dialect

Niedzielski

+++++

siehe auch Kommentare von clopper05.pdf dazu (pdf S. 5)

Die Hypothese (S. 63) ist, dass die angenommenen Eigenschaften des Sprechers die Wahrnehmung des Sprachsignals beeinflussen. (Wenn man ein Sprachsignal synthetisiert, und einem Hörer erzählt, dass der Sprecher bayerisch ist, wird der Hörer das gleiche Signal nicht auf dieselbe Weise wahrnehmen, als wenn man ihm erzählt, der Sprecher sei aus Norddeutschland). Oder auch eine der Haupthypothesen ist: 'Listeners use social information just as they use visual and other information to create or calibrate the phonological space of speakers'

Es geht hier um Canadian-raising in der kanadisch-englischen Aussprache (KE). d.h. Das Amerikanisch-Englische (AE) [aʊ] entspricht KE [ɛʊ] oder [əʊ], sodass ein KE-gesprochener 'town' [təʊn] (AE = [taʊn]) sich für den Amerikaner eher wie 'tone' anhört.

Hier würde ich mich im wesentlichen (wenn nicht sogar ausschließlich) auf Expt. 1 konzentrieren. 41 Einwohner aus Detroit hörten (a) einen Detroit-Sprecher. Sie mussten dann aus (b) einem 6-stufigen Kontinuum den Vokal auswählen, der am besten zu (a) passte. Die Endpunkte des Kontinuums waren: (token 1) sehr AE-ähnlich bis (token 6) sehr KE-ähnlich. Am meisten wählten die Hörer tokens 2-4.

Wenn den Hörern erzählt wurde, der Sprecher stammte aus Kanada, wählten sie eher Token 4 aus (mehr KE-ähnlich). Wenn den Hörern erzählt wurde, der Sprecher stammt aus Michigan/Detroit, wählten sie überwiegend Tokens 3 und 2 (mehr AE-ähnlich).

Die Hauptschlussfolgerung

Erstens. 'Detroiters expect to hear raised /aw/ in the speech of Canadians, and therefore, they do.'

Zweitens. Wenn man den Hörern erzählt, der Sprecher sei aus Detroit, dann hören sie eher sehr amerikanisch-klingende Vokale (tokens 2, 3). d.h. der Sprecher mit dem Label "Detroit" klingt für sie amerikanischer als seine Aussprache tatsächlich der Fall ist - denn eigentlich entspricht token 4 seiner Aussprache, aber dies wurde kaum ausgewählt, wenn die Hörer meinten, er sei aus Detroit. Das heißt: die stereotypischen Vorurteile (für wie ein Detroiter oder Kanadier klingen soll) beeinflussen die Wahrnehmung.

Es gab scheinbar keine Unterschiede zwischen männlichen und weiblichen Hörern (S. 69).

Die wichtigste Schlussfolgerung S. 79: 'Detroiters expect to hear raised /aw/ in the speech of Canadians, and therefore, they do.'

purnell99.jlangsocpsych.pdf

-----

### Experiment 1

Ein Sprecher (a) rief verschiedene Vermieter an und sagte am Telefon "I'm calling about the apartment you have advertised" in einem von drei (CE: Chicago Englisch; AAVE: African American Vernacular English; SAE: Standard-Amerikanisch-Englisch) Aussprachevarianten. Die zu prüfende Hypothese war: die Aussprachevarianten beeinflussen die Wahrscheinlichkeit einer Besichtigungseinladung. Genauer: die Hypothese ist, dass hier die Mehrheits-Aussprache in der Gemeinde den größten Einfluss hat. Daher wird der Anrufer eher eine Einladung bekommen, wenn er in z.B. einer CE-Variante spricht, und einen Vermieter anruft, der in einer Gemeinde lebt, wo am meisten CE gesprochen wird. S. 14 "Tables 1, 2, and 3 show that the percentage of appointments made in each locale corresponds approximately with the ethnic makeup of the geographic area. "

Auch "Where SAE is concerned (see Figure 1), we see a relatively level pattern across the locales regardless of population and house- holder density (60% to 70%)." d.h. dass die Standardaussprache überall akzeptiert wird, unabhängig von der relativen Häufigkeiten der regionalen Aussprachen.

### Experiment 2

Das zweite Experiment (kann überflogen werden) bestätigt nur, dass der Sprecher (Baugh) die 3 Aussprachevarianten mit einer guten Genauigkeit produzieren konnte.

Experiment 3: Nur das Wort *hello* ausgeschnitten aus den drei Dialekt-Sätzen des Sprechers wurde Hörern vorgespielt. Hier wieder alle Details überfliegen: es konnte gezeigt werden, dass Hörer aufgrund dieses einen Wortes zwischen den drei Dialekten unterscheiden konnten.

Expt 4: (kann eventuell komplett weggelassen werden): es gibt dialekt-spezifische akustische Unterschiede in der Produktion des Wortes *hello*.

Siehe dann bitte Kommentare und Kritik dazu in clopper05.pdf, pdf S. 8

edwards2014b.pdf

+++++

Die Studie befasst sich hauptsächlich mit der Frage, ob Kinder, die Dialekt zu Hause/ in der Gemeinde aber die Standardsprache in der Schule sprechen, größere Schwierigkeiten haben, Wörter in der Standardsprache zu erkennen.

Hier ist der Dialekt: African-American-English (AAE) und die Standardsprache Mainstream American English (MAE).

Es gab zwei Aufgaben:

Dialekt-Erkennung: inwiefern konnten die AAE-Kinder zwischen den Dialekten unterscheiden?

Dialekt-Verständnis: inwiefern war das Verständnis von Wörtern produziert in MAE komprimiert, wenn sie in Kontexten vorkamen, die zweideutig in AAE (z.B. coal/cold = / koul/ in AAE) jedoch nicht in MAE waren?

Diese Ergebnisse wurden korreliert mit (3) Dialekt-Dichte (4) die Wortschatzgröße (siehe unten).

1. Dialekt-Erkennung: Es gab 6 Sprecher, 3 AAE, 3 MAE. Ihre Stimmen wurde synchronisiert mit 'Cartoon-Monsters', die die Stimmen sprachen. Die Monsters waren farb-kodiert. Training-Phase: die Kinder sahen die sprechenden-Monster zu, (es gab 2 Monster auf dem Bildschirm in den 2 unterschiedlichen Farben) und mussten danach hindeuten, welches Monster gesprochen hat.

Testphase: das akustische Signal wurde präsentiert, ohne dass das Monster sprach. Die Kinder mussten raten, welches Monster es war. Die akustischen Signale waren von 3 zuvor gehörten Sprechern; und auch von 3 neuen Sprechern.

2. Dialekt-Verständnis: eine Bildbenennungsaufgabe. Ein Bild, und eine AAE-Stimme. wurden präsentiert. Trainingsphase: Bilder wurden in Gruppen von 3 gezeigt und ausgesprochen in AAE. Die Bilder wurden auf eine solche Weise gewählt, dass zwei Wörter zweideutig in AAE waren (coal/cold = /koul/ in AAE) aber nicht in MAE (/koul, kould/). Oder Bilder wurden gewählt mit Mehrzahlen, die in AAE zweideutig sind (d.h. Kinder sahen ein Bild von zwei Katzen, hörten die AAE-Aussprache 'two cat').

Testphase: Die Kinder sahen eine Gruppe von 3 Bildern und hörten in MAE "Show me X" wo X eins der Bilder war. d.h. die Kinder hörten in der Trainingsphase AAE aber in der Testphase MAE. Die Frage ist dann z.B. wenn sie MAE /kould/ oder 'cats' in der Testphase hörten, ob sie das Bild richtig benennen können oder nicht.

3. Dialect-Density (S. 14) wurde für jedes Kind gemessen - d.h. wie stark es Dialekt spricht. (= die Anzahl der Dialektmerkmale als Proportion der Gesamtanzahl der Wörter, die von dem Kind produziert wurden).

4. Wortschatzgröße (S. 8). Alle Kinder wurden für 'expressive vocabulary' gemessen =grob gesehen, die Wortschatzgröße in ihrer Sprachproduktion.

#### Diskussion (S. 21)

-----

Es gab eine Beziehung:

zwischen (1, 4),

jedoch nicht zwischen (1, 3)

zwischen (2, 4)

zwischen (2, 3) 'dialect density was negatively related to MAE comprehension' also je stärker der Dialekt, umso schlechter war MAE-Dialekt-Verständnis

Die Hauptschlussfolgerungen:

'better linguistic and metalinguistic skills may help children learn how to dialect shift. children who speak a more dense non-mainstream dialect may need to expend more cognitive resources simply to understand classroom discourse in MAE'

clopper05.pdf

+++++

Das zusätzlich interessante aus clopper05.pdf

1. pdf S. 13. und Abb. 13.3. Hier zeigt Clopper, dass Sprecher, die nur in einem Dialektgebiet gewohnt haben (homebodies), weniger imstande sind, einen Dialekt zu

erkennen als Sprecher (army brats), die in mindestens 3 Gebieten gewohnt haben. d.h. wenn man in verschiedenen Dialektregionen gewohnt hat, wird man eher imstande sein, Dialekte voneinander perzeptiv zu unterscheiden.

## # 17 Determinism in dialect contact

trudgill2000.jling.pdf

+++++

Es geht um Dialekt-Mischung also um den Dialekt, der entsteht, wenn Bevölkerungen mit unterschiedlichen Dialekten aufeinander stoßen, wie es der Fall im Neuseeländisch-englischen (NZE) war.

Die Haupthypothese. Die Entwicklung eines solchen Dialektes ist nicht willkürlich, sondern aus den Input-Dialekten und Populationsgrößen der Dialekte vorhersagbar. Diese Entwicklung kommt aufgrund der spontanen, gegenseitigen Imitation vor. d.h. wenn eine Gruppe von Personen, die über einem langen Zeitraum isoliert von anderen Gemeinschaften zusammenwohnt, dann entwickelt diese Gruppe allein aufgrund des gegenseitigen Kontaktes und darauf entstehenden Imitation ihren eigenen Dialekt.

Es handelt sich hier um eine auditive Analyse von 83 Sprechern des ONZE-Korpus geb. 1850-1889 aus 34 verschiedenen Gegenden in Neuseeland, 39 aus der Nord- und 44 aus der Südinsel (S301). Diese Daten wurden verwendet, um die Frage zu beantworten: warum ist phonetisch NZE so wie es ist?

Es handelt sich also um den sogenannten Koineisation /koi.ne.i.zeiʃn/ = die Entstehung eines neuen Dialektes wenn mehrere Dialekte über einem längeren Zeitraum in einer geschlossenen Gemeinschaft auf einanderstoßen; und um die Frage, ob die Merkmale von NZE aus den Input-Dialekten (aus Großbritannien und teilweise aus Australien) vorhergesagt werden können. d.h. warum gingen einige Merkmale verloren, und warum wurden andere modifiziert?

Die wichtigsten Dialekte, die zu der Entwicklung von NZE beigetragen haben, sind auf S. 303.

Welche Phasen gibt es also in der Entwicklung eines neuen Dialektes?

Rudimentary levelling: das bedeutet, dass einige Merkmale der Input-Dialekte gleich z.B. während der langen Schiffsreise aufgrund von Kontakt zwischen den Reisenden nach Neuseeland verloren gegangen sind. Die Evidenzen dafür: z.B. gab es damals in einigen südöstlichen Dialekten von England einen /v, w/ Merger (sodass /w/ den /v/ in Wörtern wie 'village' ersetzt hat) - dieser Merger findet man aber unter den älteren Leuten aus dem ONZE-Corpus nicht. Trudgill will hier argumentieren, dass der Wegfall dieses Merkmals etwas mit der gegenseitigen Imitation (während und kurz nach der Schiffsreise) zu tun hat. Es ist aber wahrscheinlicher, dass dieses /v, w/ Merger gerade ein Lautwandel-im-Fortschritt in Südostengland war, an dem die Reisenden ungenügend teilgenommen haben.

Die zweite Phase ist 'extreme variability'. Dies ist eine Zeit, in der die verschiedenen Dialekte zum großen Teil koexistieren. Evidenzen dafür ist die hohe Intra-Sprecher-Variation. Ein aus dem ONZE-Korpus untersuchter Herr Ritchie wird als Beispiel angeführt, der Merkmale verschiedener britischen Dialekte zeigt (S. 305). Dann auch andere Sprecher (S. 306), die eine sehr große Vokal-Variabilität zeigen.

Es gibt auch in dieser zweiten Phase sehr viel Variation zwischen Sprechern, (inter-individual variability) da in dieser Phase die phonetischen Merkmale eines Individuums sogar innerhalb derselben Familie sehr idiosynkratisch sind (Beispiel Ritchie S. 306).

Die dritte Phase, focusing (S. 307). Das ist die Phase, in der sich der neue Dialekt gestaltet, und in der die Intra- und Interprechervariabilität reduziert wird. Vielleicht dann zu S. 309 springen.

Der Grund weshalb NZE viele Merkmale des Südostenglischen hat, ist nicht weil Leute aus dieser Gegend in der Mehrheit waren, sondern eher weil das Südostenglische viele Merkmale mit den anderen Varietäten gemeinsam hatte, die in Neuseeland vertreten waren. d.h. es sind die phonetischen Merkmale aus der südostenglischen Varietät, die es auch in den anderen Varietät gab, die im NZE überlebt haben.

Dagegen die phonetischen Merkmale aus der südostenglischen Varietät, die eigenartig waren - die es nicht in den anderen Varietäten gab - sind eher im modernen NZE verloren gegangen. Insbesondere die Tendenz im südostenglischen /h/ am Wortanfang wegzulassen ('hammer' ohne /h/) kommt im NZE nicht vor, weil das auch kein Merkmal z.B. des Schottisch- oder Irisch-Englischen war.

Zu unmarked Varianten (S. 311). Laut Trudgill sind es 'unmarked Varianten' die bei einer Dialektmischung überleben werden.

Obwohl die meisten Besiedler nach Neuseeland damals *roses* von *Rosa's* /ɪ, ə/ differenzierten, überlebte dieser Kontrast in NZE nicht (daher sind u.a. 'dances'/'dancers' oder 'boxes'/'boxers' sowohl in NZE als auch in AusEngl homophon: beide Wörter der Paare haben /ə/). Trudgills Erklärung: weil /ə/ unmarkiert ist.

Unsere Erklärung (Harrington & Schiel, 2017). Weil /ɪ/ akustisch in die Richtung von /ə/ streut. Siehe <http://www.phonetik.uni-muenchen.de/~jmh/papers/harringtonschiel.pdf>, pdf Seite 431, 'Such a model could begin to explain why New Zealand English (NZE) neutralizes...!'

Weitere Beweise für Trudgills Theorie (S. 312)

- die Interdentalen [θ, ð] verschwanden, weil sie nur für eine Minderheit (Irisch-Englisch) typisch waren.
- /a/ in START/BATH ist nah an KV 4 in Neuseeland-Englisch; eventuell weil ein sehr vorderer [a] am häufigsten in den Britisch-Inseln für solche Wörter damals im 19 Jahrhundert produziert wurde.
- /ɪ, ε/ waren im frühen Neuseeland-Englisch phonetisch sehr hoch. /ɪ/ wurde dann später (ab ca. 1950) zentralisiert (Folie anbei). Die Argumentation, weshalb /ɪ, ε/ in der sich entwickelnden Aussprache von Neuseeland-englisch höher waren (S. 312), ist nicht besonders überzeugend.

Diphthong-Shift: [ai] -> [oi], [au] -> [ɛu]

In ausgeprägteren Australisch- und Neuseeland-Englisch Varietäten ist heutzutage:

- [ai] (buy) oft wie [oi]
- [au] (down) oft wie [ɛu]
- 

Das ist aus diesem Grund:

63 of our speakers (75%) have diphthong shifting of at least some diphthongs.

NB: Diphthong-Shift ist/war vor allem typisch für London-Cockney-Englisch

Bezüglich 'fir', 'fur', 'fern'

Die Argumente werden jetzt (S. 314) etwas schwach. Dieser Kontrast (also unterschiedliche Vokale) gab es für mehrere Sprecher, die im 19. Jahrhundert damals nach Neuseeland unterwegs waren, aber heutzutage hat NZE /ɜ/ (langer zentralisierter /ə/). Das ist weil Mergers so oft im Lautwandel vorkommen.

Die Argumentation für Lot eventuell weglassen.

Die Argumentation für /a/ vor Nasalen ('dance') und vor Frikativen ('laugh') usw. Im 18. Jahrhundert in englischen Varietäten waren diese Vokale kurz, sodass 'ant' und 'aunt' beide /ænt/ waren. 'laugh' war auch /læf/ mit kurzem /æ/ (nah an KV 4). Das ist heute noch so in den meisten nordenglischen Varietäten, sowie im Schottisch-englischen als auch im Irisch-englischen: keine dieser Varietäten differenziert ant/aunt.

Dann im ca. 18. J. in Südost England und in der Standardsprache wurde dieser kurze Vokal vor einigen Nasalen und Frikativen lang und auch rückverlagert d.h. von einem KV4 zu einem langen KV5 also von /æ/ zu /ɑ:/ in 'aunt' und 'bath' und 'laugh'

Das war ein Lautwandel-im-Fortschritt gerade in der Zeit (18-19. Jahrhundert) der Besiedlung in Australien und Neuseeland. Die Australier und Neuseeländer haben damals mehr oder weniger an diesem Lautwandel-im-Fortschritt teilgenommen haben: daher haben sie heutzutage einen langen Vokal in 'dance' und 'aunt' (und differenzieren daher auch ant/aunt). Ein wesentlicher Unterschied im Vgl. zum englischen Standard ist, dass im Australisch- und Neuseeländisch-englischem dieser lange Vokal nicht KV5 wie im Standardenglischem sondern \*zentral\* als zwischen KV4 und KV5 ist - ähnlich aber phonetisch etwas weiter nach vorne wie der deutsche /a:/ in 'Lahm'.

Eventuell gab es (das ist meine Interpretation) zwei Lautwandel: Längung und Rückverlagerung also /æ/ -> /æ:/ -> /ɑ:/. Die Australier und Neuseeländer haben teilgenommen an /æ/ -> /æ:/ aber nur teilweise an /æ:/ -> /ɑ:/. Oder - hier wäre noch eine Idee. Der Vokal ist nicht komplett rückverlagert wie im Standardenglischen von heute, sondern zentralisiert aufgrund der gegenseitigen Imitation während und nach der Schiffsreise mit vielen Sprechern, die einen vorderen /æ/-Vokal hatten - d.h. der Kontakt zu diesen Leuten mit /æ/ hat die Rückverlagerung gebremst.

Post-vokalisches /r/.

Hier wird Trudgills Argumentation wieder sehr schwach.

Die meisten Sprecher in dem ONZE-Corpus - und die meisten die damals nach Neuseeland im 19. Jahrhundert ausgewandert waren - hatten ein /r/ in Wörtern wie 'car', 'bird' usw. also wo ein 'r' in der Orthographie vorkommt.

Aber NZE ist heute non-rhotic. Wie Trudgill sagt: 'We cannot explain the loss of post-vocalic 'r' by any of the mechanisms so far described'.

trudgill08.langinsoc.pdf

+++++

Die Frage hier ist, ob sich in neu besiedelten Kolonialgebieten - wie Neuseeland - die Aussprache aus deterministischen Gründen erfolgt ( also aufgrund der gegenseitigen Imitation), oder ob sich eine identifizierbare Aussprache aufgrund von Identität entwickelte - d.h. weil die Besiedler eine Nationalidentität getrennt von Nationalitäten anderer Länder etablieren wollten.

Nicht alle Forscher (S242) folgen dieser Idee von Dialekt-Mixture, auch wenn sie nicht unbedingt die Theorie der Identität unterstützen. d.h. bis Trudgill ist diese Idee, dass sich ein neuer Dialekt deterministisch aus den Input-Dialekten entwickelte, kaum akzeptiert worden.

S 243: das ist ganz wichtig. Trudgill sagt, dass Dialekt-Mischung deterministisch zustande kommt, weil wir uns gegenseitig imitieren. Jetzt vorspringen zu S 252, Imitation: 'linguistic accommodation is not driven by social factors. Linguistic diffusion and new-dialect formation are "mechanical and inevitable" because linguistic accommodation is automatic, because, as Cappella (1997:69) says, it is an aspect of "the relatively automatic behaviors manifested during social interaction... My suggestion is therefore that it is this innate tendency to behavioral coordination, not identity, that is the very powerful drive that makes dialect mixture an almost inevitable consequence of dialect contact'.

Und auch Labov (2001:506), S. 251:

'every speaker is constantly adapting his speech-habits to those of his interlocutors . . . The inhabitants of a settlement . . . talk much more to each other than to persons who live elsewhere.... (S. 252): The principle of density implicitly asserts that we do not have to search for a motivating force behind the diffusion of linguistic change. The effect is mechanical and inevitable; the implicit assumption is that social evaluation and attitudes play a minor role.' (Labov 2001:20)

d.h. die Art der Imitation, die zum Dialekt-Mixture führt, ist automatisch: ein Sprecher imitiert nicht in dieser Situation einen anderen weil er die Person oder seine Aussprache mag. Imitation ist statt dessen biologisch/motorisch (nicht linguistisch) und kommt zustande, auch weil wir Körperbewegungen usw. gegenseitig und spontan imitieren.

S. 243: Tuten vertritt jedoch die Idee, dass Imitation (accommodation) vor allem in solchen neu besiedelten Kolonialgebieten wie in Neuseeland zustande kommt, weil die Bevölkerung durch die Merkmale der Aussprache eine nationale Identität ausdrücken wollte.

Trudgill versucht dann verschiedene Beispiele zu bringen, um die Hypothese von Dialekt-Mixture zu unterstützen. Ich würde das überfliegen und nicht alle Beispiele besprechen, sondern eventuell nur diese 3.

1. Die Entwicklung von Irisch-Englisch in den 14-15 Jahrhunderten sei eine Mischung der Dialekte aus verschiedenen westlichen- und südwest englischen Varietäten. Hier

spricht Trudgill auch von der Möglichkeit, dass gemittelte phonetische Varianten durch Mixture entstehen könnten. zB

'each' in der westmidlands Varietät im 15 Jahrhundert war [ux] und in der südwestlichen Varietät war 'each' [ex]. Die phonetische Mischung daraus, die es dann im Irisch-englischen gab war [eux] oder vielleicht etwas wie [œx] also eine phonetische Form mit Merkmalen aus beiden Dialekten, die Input zum Mixture waren.

2. zB die romanischen Varietäten in Valencia im 13 Jahrhundert waren:

Ostkatalanisch mit /e, ε, ə/

Aragonese mit /e, iə/

Westkatalanisch mit /e, ε/

Dialekt-Mixture würde vorhersagen, dass die Vokale, die diese 3 Dialekt \*nicht\* gemeinsam haben, mit der Zeit wegfallen müssten; und genau dies ist geschehen, sodass die Aussprache in Valencia nur einen Kontrast /e, ε/ hat.

3. Altenglisch (OE) war sehr variabel, zumindest morphologisch - eventuell weil das die zweite Phase des Dialekt-Mixtures ist (in dem alles koexistiert). Daher gab es gleichzeitig scheinbar zB 2 Formen von 'ich bin', 'du bist' ('eom'/'eart' aus dem Altnordischen/Gothischen, die andere 'beom'/'bist' aus dem Altsächsischen und Althochdeutsch).

Conclusion

-----

Die wichtigste Schlussfolgerung:

Identity is not a powerful enough driving force to account for the emergence of new, mixed dialects by accommodation... identity factors cannot lead to the development of new linguistic features, and it would be ludicrous to suggest that Australian English speakers deliberately developed, say, wide diphthongs in order to symbolize some kind of local or national Australian identity.

holmes08.langinsoc.pdf

+++++

Dies ist ein Kommentar zu trudgill08.langinsoc.pdf

Diese Soziolinguisten (die sich eventuell durch Trudgills Idee angegriffen fühlen!) wollen argumentieren, dass diese mechanistische, deterministische Idee der Dialekt-Mischung nicht ausreicht, um die phonetische Form des Dialektes vorherzusagen, wenn Dialekte aufeinander stoßen.

Sie vertreten eher die Idee, dass es mehrere soziale Faktoren geben muss, die die sich entwickelnde Aussprache prägt. Der wesentliche Unterschied zwischen Holmes & Kerswill auf der einen Seite und Trudgill auf der anderen:

Trudgill... identifies "accommodation in face-to-face interaction" as "the fundamental mechanism leading to dialect mixture." By this it subsequently emerges he means a "biologically given drive to behave as one's peers do." Thus Trudgill favors biological rather than attitudinal and social factors in resolving the dilemma of who accommodates to whom.

d.h. Trudgills Idee ist, dass accommodation rein biologisch determiniert wird (die linguistische Anpassung ist eine Folge der Tendenz, sich in der Motorik und Gestik gegenseitig zu imitieren); Holmes & Kerswill dagegen: unsere Einstellung zu anderen Leuten und zu anderen sozialen Merkmalen (ob wir zB eine soziale Schicht mögen oder nicht) sind genauso wichtig/ eventuell noch wichtiger.

Holmes & Trudgill fragen:

Why would one person accommodate to another's speech rather than the other way around?

Unsere (Harrington/Schiel) Antwort dazu: das passiert wenn die phonetischen Varianten einer Person in die Richtung einer anderen streuen, aber nicht umgekehrt -

<http://www.phonetik.uni-muenchen.de/~jmh/papers/harringtonschiel.pdf>  
Abb. 1 pdf S. 48 (oder S. 421).

Das sind hypothetische phonetische Varianten in einem 2D-Raum von zwei Sprechern unterschiedlicher Dialekte. Unsere Hypothese: B verschiebt sich Richtung A, weil B sich in Richtung A streut, aber nicht umgekehrt.

Holmes & Kerswill: It seems sociolinguistically axiomatic that a desire to be like the other person in a range of possible social respects (status, social group, social role, etc.) is one fundamental explanation.

Sie führen - soviel ich sehen kann - keine aktuellen Beweise, um ihre Argumentation zu unterstützen. Sie haben sicherlich recht jedoch, dass man kaum gegen Trudgill argumentieren kann, weil er seine Theorie anhand von historischen Beispielen veranschaulicht, die durch die gesprochene Sprache zum Teil nicht nachgewiesen, oder falsifiziert werden können.

trudgill08b.langinsoc.pdf

+++++

Dies ist ein Kommentar von Trudgill an diejenigen gerichtet, die ein Kommentar zu trudgill08.langinsoc.pdf geschrieben hatten, darunter holmes08.langinsoc.pdf

Die wichtigsten Punkte seiner Antwort:

A number of the commentators have alluded to a link between identity and accommodation; but normal coordinatory behavior, including linguistic accommodation, is so automatic – children do not operate it “at will,” as Holmes and Kerswill believe.

Er unterstützt seine Argumentation wie folgt.

1. Wenn Dialekte aufeinanderstoßen, und wenn vor allem Kinder dieser unterschiedlichen Dialekte miteinander interagieren, ist das Ergebnis IMMER irgendeine Dialektmischung, die die Merkmale beider Dialekte hat, oder die diese Merkmale mittelt.

2. Soziale Faktoren spielen für Trudgill keine Rolle: the initial “work” of new-dialect formation in these situations is performed by the small children of the first two generations, which is why the social factors which Holmes and Kerswill would like to look for do not come into play. The adult settlers in any colony, of course, bring all the

baggage of “the social marking of linguistic variants” But this kind of baggage is not relevant to 7-year-old children in the colonial situation – which is precisely why the mixing of variants from different dialects of different regional origin and different degrees of social status, with a mass of different associations and connotations, always takes place.

## # 18. Intrinsic and extrinsic vowel normalisation

adank04.jasa.pdf

=====

Die Autoren verwenden verschiedene Algorithmen, um festzustellen, ob man sprecher-normalisieren kann, ohne die phonetischen Unterschiede zwischen Sprachen oder Dialekten zu zerstören. Die Daten waren von 160 Sprechern aus zwei niederländisch-sprechenden Regionen.

Zu den Algorithmen: Die Formel Hz in Bark oder Hz in Erb braucht man nicht erwähnen. Das sind logarithmische Transformationen, die ein Verhältnis zur Wahrnehmung haben.

Dann die Algorithmen:

Syrdal & Gopal:

Das ist in einer  $F3b - F2b$  (y-Achse) x  $F1b - f0b$  (x-Achse)-Ebene. Die Grundlage für  $F3$  und  $f0$  siehe auch:

<http://www.phonetik.uni-muenchen.de/~jmh/papers/acoustics2.pdf>

S. 85/86

Gerstman: siehe harrington99.pdf.

Pro Sprecher und pro Formant zieht man das Formant-Minimum ab und dividiert durch den Formant-Bereich (= die Variationsbreite von min zu max). Z.B. wenn der Vokal mit  $F2$ -Minimum = 400 Hz und mit  $F2$ -Maximum = 2800 Hz, und  $F$  ist der Formant-Wert von irgendeinem Vokal vom selben Sprecher, dann pro Vokal:  $(F - 400)/2400$

Lobanov: Standardnormalisierung. Pro Sprecher vom Formantwert den Mittelwert abziehen und durch die Standardabweichung (desselben Sprechers) dividieren.

Nearey: schnell überfliegen und eventuell harrington99.pdf S. 78 zusammenfassen.

Nordström: hier versucht er, die Unterschiede in der Länge des Vokaltrakts zwischen Männern und Frauen zu kompensieren, indem er  $k = F3m/F3w$  berechnet ( $F3m$  oder  $F3b =$  Mittelwert über alle männlichen oder weiblichen Versuchspersonen). Und multipliziert dann jeden Formant einer Frau mit diesem  $k$  (die männlichen Formanten werden nicht verändert, wenn ich das richtig verstanden habe).

Miller (basiert auf formant-ratio Theorien, siehe z.B. harrington99.pdf S. 75 und auch die Diskussion dazu in johnson(2).pdf unten). Die Formel 12-15 sind eigentlich ganz klar - die erste basiert auf dem geometrischen Mittelwert von  $f0$  eines Sprechers.

Geometrischer Mittelwert: Mittelwert, der als die  $n$ -te Wurzel aus dem Produkt der  $n$  beachteten Zahlen berechnet ist

z.B. geometrischer Mittelwert von 2, 4, 8 =  $(2 * 4 * 8)^{(1/3)}$ .

Wenn Ihr mit (12-15) nicht klar kommt, melden.

Abschnitt A von dem Paper.

.....

Es wird eine Klassifikation durchgeführt (Results III). D.h. Wahrscheinlichkeitsmodelle werden berechnet, einmal pro Vokalkategorie. Und dann wird die Zugehörigkeit pro Vokal-Token berechnet. Was rauskommt, ist eine Verwechslungsmatrix.

z.B. in R:

```
library(MASS)
pfadu = "http://www.phonetik.uni-muenchen.de/~jmh/lehre/Rdf"
v.df = read.table(file.path(pfadu, "v.df.txt"))
# Wahrscheinlichkeitsmodelle berechnen
v.qda = with(v.df, qda(cbind(F1, F2), V))
# Klassifikation
erg = with(v.df, predict(v.qda)$class)
tab = table(v.df$V, erg)
tab
erg
```

```
  a E I
a 122 3 1
E 10 62 10
I 1 5 164
```

d.h. 122 /a/s wurden richtig als /a/ erkannt; 62 /E/s als /E/. Die Klassifikationsrate = Summe der Diagonale/Anzahl der Vokalen:

```
sum(diag(tab))/sum(tab)
```

0.92 oder 92% Auf eine ähnliche Weise sind die Zahlen in Table II berechnet worden. Es gibt laut Tabelle II die höchste Klassifikationsrate, nachdem man die Vokale Lobanov-normalisiert hat.

Abschnitt B von dem Paper:

.....

Das gleiche wie A, aber hier klassifizieren sie nicht für Vokale sondern für Gender (d.h. 93% in Tab. III in Hz bedeutet: das Hz-System konnte bei 93% der Vokale erkennen, ob sie von einem Mann oder Frau produziert wurden).

Das Ziel hier ist auf 50% herunterzukommen (weil es nach einer perfekten Normalisierung zwischen Männern und Frauen keine erkennbaren Unterschiede in den Vokalen mehr geben soll). Hier schneidet wieder Lobanov am besten ab.

Abschnitt C von dem Paper

.....

Diesmal wurde für Region klassifiziert. Es gibt 8 Regionen, daher ist die Zufallsrate 12.5%. Man bevorzugt eine Normalisierung, in der die soziolinguistische Information erhalten bleibt (weil das Ziel einer Normalisierung sein soll, die anatomisch bedingten Unterschiede zu entfernen, nicht aber die linguistischen/soziolinguistischen).

Tabelle IV zeigt wie erfolgreich die Dialekte erkannt werden konnten (getrennt aufgeteilt pro Vokal).

Ohne Normalisierung kann man Dialekt richtig in ca. 30% der Fälle erkennen (z.B. Hz 30%). Mit Normalisierung soll die Klassifikationsrate so wenig wie möglich unter 30%

fallen. Daher zeigt hier Syrdal & Gopal mit 25% am wenigsten Erfolg; und Nearey und Nordström den größten (30%).

#### Abschnitt D

.....

Hier werden die erfolgreichsten Verfahren (Lobanov, Nearey, Gerstmann) ausgewählt, um zu sehen, wieviel Variation innerhalb eines Faktors nach der Normalisierung übrig bleibt.

z.B. für Faktor Vokal soll die übrig bleibende Varianz (sehr) hoch sein, weil unterschiedliche Vokale sich immer noch nach der Normalisierung stark von einander abweichen sollen. Ebenso für die soziolinguistische Information. Jedoch soll die Variation wegen Gender gering gehalten werden, weil Gender nach einer optimalen Normalisierung nicht mehr differenziert werden soll.

Die Schlussfolgerung: Extrinsische Algorithmen sind am erfolgreichsten.

vorperian07.jshlr.pdf

+++++

Das Paper befasst sich mit Formantwerten von Kindern aus 14 verschiedenen Studien. Alle Amerikanisches Englisch, aber unterschiedliche Dialekte (S. 1514 Spalte 2).

Auf Anhieb dachte ich, das Paper sei ganz interessant. Aber wenn man sich das näher anschaut, hat die Argumentation viele Lücken und ist teilweise schlecht durchdacht. Das Problem: sie führen eine Reihe von Bildern an, erklären aber ungenügend oder auf eine verwirrende Weise was dahintersteckt.

Noch ein Problem: sie stellen Mittelwerte dar, und keine Varianzen.

Ich würde daher dieses Paper sehr knapp zusammenfassen, und nur 2 Sachen besprechen, nämlich:

1. Altersbedingte Änderungen der Formanten - also: wie nähern Formanten eines Kindes sich während der Entwicklung an den Formanten eines Erwachsenen an?
2. Interaktion von 1. mit Geschlecht (und mit Vokal) - also ab welchem Alter merkt man geschlechtsspezifische Unterschiede in den Formanten? Und ist das für alle Vokale so? (= Interaktion mit Vokal).

Jetzt bitte gleich zu den Hauptergebnissen springen, die als a-h in der zweiten Spalte unter General Discussion auf S. 1530 gelistet sind.

Welche sollten besprochen werden? Hier ist meine Meinung dazu:

(a) Die Vokale von Kindern - sogar im ganz frühen Alter - können so wie jene von Erwachsenen im F1 x F2 Raum dargestellt werden. Die Aussage ist trivial. Vergessen.

(b) Mit zunehmendem Alter senken sich die Formanten.

Das ist so, weil der Vokaltrakt mit zunehmenden Alter länger wird. (S. 1511 'Vocal tract length in neonates is about 6–8 cm, compared with an average length in adult females of about 15 cm and in adult males of about 18 cm').

Vielleicht dies belegen durch Abb. 3 und 4 auf derselben Folie nebeneinander.

(c) Die Variabilität der Formanten wird geringer mit zunehmenden Alter insbesondere im F1.

Das ist so, weil sich die phonetische Variabilität mit der Zeit stabilisiert, aber so viel ich sehen kann, gibt es wenig Beweise dazu im Paper. Eventuell weglassen.

(d) Männlich-weiblich Unterschiede im Vokalraum zeigen sich schon ab dem 4ten Lebensjahr.

Das ist vielleicht ganz interessant. Diese Aussage eventuell mit Abbildungen 9 und 10 beleuchten aus Perry (siehe auch die Abbildungen in johnson.pdf/johnson2.pdf). Man sieht hier m-w Unterschiede schon im Alter von 4 Jahren. Insbesondere ist F1 niedriger für Jungen als für Mädchen. Dieser Unterschied wird progressiv größer mit zunehmenden Alter.

Der Grund für die geschlechtsspezifischen Unterschiede ist ca. wie folgt.:

In erwachsenen Männern ist das Verhältnis von Rachenraum zu Mundraum größer als für Frauen (S. 1521). Das haben die Autoren richtig erfasst. Aber was in dem Artikel nicht so deutlich steht: die kleinere Länge vom Rachenraum in Frauen führt im wesentlichen zu einer Hebung von F1 in offenen Vokalen wie [a] und zu einer Hebung von F2 in hohen vorderen Vokalen wie [i]. Das ist schon alles von Fant (1973) <https://mitpress.mit.edu/books/speech-sounds-and-features> festgestellt worden. Das heißt also: Vokal und Geschlecht interagieren miteinander. Das sieht man vor allem in den Erwachsenen in Fig. 11, in der ganz klar gezeigt wird, dass die m-w Unterschiede am kleinsten in [u], und viel größer in [i, a] sind.

Dann ist die Frage, inwiefern man ähnliche Muster bei den Kindern sieht.

Dazu dann fig. 9. In jedem Fall ist der m-w Unterschied sehr groß in [a], wie für die Erwachsenen auch. Der m-w Unterschied wird progressiv (mit zunehmenden Alter) größer in [i]. Aber der m-w Unterschied wird *auch* progressiv größer in [u] - und das passt nicht zu dem Muster der Erwachsenen, und auch nicht was man aus der Physiologie (des größeren Rachenraums in Männern) erwarten würde. Das deutet darauf hin, dass die m-w Unterschiede im Alter von 12 und 16 in [u] wegen Aussprache/Dialekt zustande kommen (siehe johnson.pdf/johnson2.pdf dazu) - also dass die jungen Frauen ein *phonetisch* vorverlagertes [u] haben. Das hat nichts mit Anatomie zu tun, sondern ist wahrscheinlich soziophonetisch bedingt. Wir wissen auch, dass [u] für Aussprache/Dialektunterschiede recht anfällig ist. All das steht nicht in dem Artikel. Aber eine andere Interpretation kann es nicht geben. Physiologisch bedingt sind diese [u]-Unterschiede zwischen Jungen und Mädchen ganz sicherlich nicht. Die Autoren kommen aber doch zu einer ähnlichen Schlussfolgerung: 'The increased variability of F2 for the high-back vowel /u/ across the entire developmental age range may indicate that variability is rooted in several factors, including the influence of dialect'

(e) Die Änderungen in den Formanten sind nicht-linear, d.h. es gibt plötzliche Formantänderungen, die mit einem schnellen Zuwachs in den Vokalorganen oder Teilen davon verbunden sind.

Dies eventuell durch Abb. 2 veranschaulichen: 'In Figure 2, summary of male acoustic data, there is a noticeable overall jump in the F1–F2 vowel acoustic space between the ages of 14 and 15 years when abrupt drops in F1 and F2 formant frequencies can be noted for all corner vowels.' (S. 1520). Dies ist so, weil sich bei Jungen im Jugendalter/ Pubertät der Kehlkopf noch einmal senkt.

Ebenso durch Abb. 4: 'there is a noticeable overall jump in F1–F2 vowel acoustic space between the ages 1 and 4 years, that is, an abrupt change in the first and second formant frequencies for all corner vowels.

Dies ist so, weil sich der Kehlkopf in der frühen Kindheit senkt (wodurch der Vokaltrakt plötzlich länger wird).

(f)  $f_0$  reduziert sich sogar schon nach dem ersten Geburtsjahr. Die Reduzierungen sind am schnellsten zwischen 1-3 Jahren und im Jugendalter.

Es gibt hierzu zu wenig Daten in diesem Paper, und wir sollten uns sowieso eher auf Formanten konzentrieren. Daher weglassen.

(g) Das velopharyngeale System wird erst ab dem ersten Lebensjahr reif genug, um orale (nicht nasale) Vokale zu produzieren.

Das gleiche wie (f). Zu wenig Daten, weglassen.

(h) Die geschlechtsspezifischen Änderungen in den Formanten kommen aufgrund von Räumen (insbesondere der unterschiedlichen Verhältnisse in der Länge vom Rachen- und Mundraum zwischen den Geschlechtern) und nicht alleine aufgrund von Vokaltraktlänge zustande.

Viel zu vage, und geschlechtsspezifische Unterschiede sind sowieso durch (d) abgedeckt.

Alles andere im Paper kann weggelassen werden.

johnson.pdf (johnson2.pdf)

+++++

Wir wollen aus Johnson feststellen, inwiefern Hörer für Sprecherunterschiede normalisieren. Das Problem an sich ist ganz deutlich in Abb. 15.2 von Peterson & Barney (S. 364) zu sehen. Die Vokale variieren sehr stark von Sprecher zu Sprecher, und insbesondere zwischen Männern und Frauen, und zwischen Erwachsenen und Kindern. Zusätzlich ändern solche sprecherspezifischen Merkmale oft dieselben Cues, die für phonetische Unterschiede benötigt werden (z.B. trennt ein hoher  $F_2$ [i] von [e], aber Frauen haben höhere  $F_2$  als Männer, sodass  $F_2$ -Mann-[i] ähnlich sein kann wie  $F_2$ -Frau-[e]. Daher: wie normalisiert der Hörer für eine solche sprecherspezifische Variation?

Einiges in diesem Artikel ist etwas durcheinander.

Vielleicht daher nur/hauptsächlich aus diesem Johnson-Artikel nehmen, was für unsere Normalisierungsfrage relevant ist. Hier eventuell die wichtigsten Sachen.

## Formant-ratio theory

Die Idee hinter dem Formant-Ratio-Theorie ist, dass wir nicht absolute Formantwerte hören, sondern das Verhältnis (Ratio) zwischen Formanten. Insofern entsteht durch die Berechnung von Formant-Ratios eine gewisse Normalisierung. Denn, wenn eine Frau für [i]  $F1 = 250$  und  $F2 = 3000$  hat, und ein Mann für den selben Vokal  $F1 = 200$  Hz und  $F2 = 2400$  Hz hat, dann ist das Verhältnis  $F2/F1$  für beide gleich.

Einige der Formant-Ratio-Algorithmen, die oben auf S. 368 gelistet werden, werden dann ausführlicher in Adank besprochen.

Ein großes Problem mit allen Formant-Ratio-Theorien ist wie folgt. Der Unterschied z.B. zwischen Männern und Frauen ist nicht linear (siehe Diskussion dazu in Vorperian oben (e)) d.h. der männlich-weiblich Unterschied in den Formanten für [a] ist größer als für [u]. Es reicht daher nicht aus, zu behaupten - wie es Formant-Ratio-Theorien tun - dass der Formant-Unterschied zwischen zwei Sprechern a, b immer ein Verhältnis  $a/b$  ist; das Verhältnis wird also sehr stark von der Identität des Vokals beeinflusst.

Formant-ratio Normalisierung ist ein sogenanntes intrinsisches Verfahren, denn man benötigt nur den einen Vokal vom Sprecher, um die Normalisierung durchzuführen. Ein Bestandteil von Formant-Ratio-Theorien ist daher, dass der Hörer auch nicht mehr als den aktuell wahrgenommenen Vokal benötigt, um die Normalisierung durchzuführen.

## Extrinsische Normalisierung

+++++

Viele andere Normalisierungsalgorithmen sind extrinsisch - man benötigt eine Stichprobe der Vokale von einem Sprecher (und die Stichprobe soll/muss die Eckvokale des Sprechers beinhalten), um zu normalisieren. Extrinsische Verfahren wie Lobanov (1971) sind oft am wirkungsvollsten, um die Streuung, die wegen Sprecherunterschieden zustande kommt, zu reduzieren. Die Frage bleibt: gibt es irgendwelche Beweise für den Bereich menschlicher Kognition, dass Normalisierung extrinsisch ist (= ein Hörer benötigt eine Stichprobe der Vokale des Sprechers, um die sprecherspezifischen Aspekte der Vokale wegzunormalisieren)? Dazu dann etwas mehr in 15.3.5 auf S.371 und das berühmte Experiment von Ladefoged & Broadbent (1957). In diesem Experiment haben Hörer einen Satz mit einem finalen 'bit' oder 'bet' wahrgenommen (also Trägersatz gefolgt von  $bVt$ ,  $V = /i/$  oder  $/\varepsilon/$ ). Die Formanten vom Trägersatz wurden künstlich/synthetisch manipuliert, in dem sie entweder nach oben oder nach unten in der Frequenz verschoben wurden. Höhere Formanten im Trägersatz führten zu mehr 'bet' Urteilen. Das ist so, weil die Formanten von bit/bet im Verhältnis zu den Formanten im Trägersatz wahrgenommen werden. Wenn im Trägersatz die Formanten generell höher sind, dann gibt es eine größere Wahrscheinlichkeit, dass der gleiche  $bVt$  Stimulus eher als /bet/ (im Gegensatz zu /bit/) wahrgenommen wird. Das wichtigste aus diesem Experiment: Sprechernormalisierung scheint also extrinsisch zu sein, d.h. die phonetische Qualität des Vokals wird im Verhältnis zu anderen Vokalen desselben Sprechers beurteilt.

Es gibt zusätzliche Beweise (Abschnitt 15.3.6), dass Vokal-Normalisierung extrinsisch sein könnte. Hörer reagieren langsamer, Vokale in einer Liste von Wörtern zu

identifizieren, wenn diese Wörter von unterschiedlichen Sprechern gesprochen wurden, als wenn sie vom selben Sprecher stammen.

#### Hörer-Erwartungen

+++++

beeinflussen (15.3.7), wie man eine Stimme kategorisiert. Hier eventuell die Studie von Johnson et al (1999) erwähnen. Hörer klassifizierten ein Kontinuum zwischen hood und hud (hʊd - hʌd), je nach dem ob den Hörern erzählt wurde, die Stimme sei von einem Mann oder von einer Frau (die Stimme war synthetisch und bezüglich des Geschlechtes ambig).

#### Acquisition of gender differences (15.5.2)

+++++

Dieses Material ist für die Vorperian-Studie von Relevanz. Es geht hier darum, inwiefern die m-w Unterschiede in von Kindern gesprochenen Vokalen wahrnehmbar sind. Einige Ergebnisse in Tabelle 15.2. Die Abbildungen in 15.4 sind ebenfalls Ergänzungsmaterial zu Vorperian.

#### Geschlechtsspezifische Unterschiede und Sprache

+++++

Hier vielleicht auf Abb. 15.5 konzentrieren, und Abb. 15.6 vergessen.

Wenn der m-w Unterschied zwischen Formanten nur anatomisch bedingt wäre, dann müssten solche Unterschiede in verschiedenen Sprachen ähnlich sein (es sei denn, man möchte argumentieren, dass der Unterschied zwischen Männern und Frauen in der Länge und Größe des Vokaltrakts für z.B. polnische im Vgl. zu schwedischen Sprechern anders ist - was also nicht sehr wahrscheinlich ist). Abb. 15.5 zeigt viel größere Unterschiede in den Formanten zwischen Männern und Frauen z.B. für Russisch im Vgl. zu Dänisch. Das bedeutet also, dass man im Spracherwerb lernt, männlich oder weiblich zu klingen - und wie genau das umgesetzt wird, hängt von der Sprache ab.

### # 19. Vowel chain shifting

Im Allgemeinen: labov94.pdf nur teilweise besprechen; samuels.pdf: eventuell nur die 1-2 Beispiele in der Diskussion von labov94.pdf berücksichtigen, wie unten beschrieben (d.h. samuels.pdf wahrscheinlich nicht näher kommentieren). Dann weiter mit der Gewichtung: maclagan > watt (d.h. es gibt viel mehr Interessantes in maclagan als in watt zu besprechen).

Aber wie immer: das sind nur Vorschläge!

labov94.pdf (und samuels.pdf)

+++++

Im wesentlichen die Prinzipien und die Patterns besprechen, und keineswegs auf die vielen Beispiele eingehen, sondern nur das wichtigste daraus entnehmen (und sowieso das Material ab. S 133 eventuell vergessen).

Was ist chain-shifting?

.....

labov94.pdf: ein Vokal (A) verschiebt sich in Richtung eines anderen Vokals (B). Vokal B verschiebt sich deshalb auch weg (und so bleibt der Kontrast zwischen den Vokalen erhalten).

Die ALternative zu einer solchen Kettenverschiebung wäre - wie Labov auf S. 119 sagt - ein sogenannter Merger (= wenn der phonologische Kontrast zwischen A und B verloren geht).

Dann vielleicht samuels.pdf S. 4 zu dem Unterschied zwischen pull-chain und drag-chain näher behandeln (S. 3-4), eventuell zusammen mit der funktionalen Erklärung von chain-shifting:

Die funktionale Theorie von Chain Shifting

.....

Dies ist eigentlich teilweise ein Vorreiter von Lindbloms H&H Theorie (und wird von Martinet, 1955, vertreten).

Der Sprecher reduziert/zentralisiert Vokale in der Spontansprache.

Diese synchrone Verschiebung könnte Lautwandel zur Folge haben.

Aber die diachrone Verschiebung von einem Vokal kann eine Kettenreaktion auf andere Vokale auslösen. Die Kettenreaktion kommt zustande, damit die phonologischen Kontraste erhalten bleiben (also damit die Bedeutungen/Semantik nicht verloren geht). Man sieht auch wegen der Kettenreaktion eine Symmetrie (laut dieser *funktionalen* Erklärung) im Vokalsystem - z.B. genau so viele vordere wie hintere Vokale; gepaarte nicht-gerundete und gerundete Vokale usw.

Siehe hier auch samuels S. 1-8.

Labov bevorzugt eine nicht-funktionale Erklärung, nämlich eher eine, die auf der Basis von Missverständnissen beruht. Dies wird teilweise in samuels S. 19 erklärt. In Abb. 2.5.2 gibt es ein Ausreißer-Token von /o/ (d.h. ein Sprecher plante /o/, aber die Formanten waren eher im /a/-Bereich), sodass das Ausreißer-/o/-Token in die Verteilung von /a/ fällt. Dieses Token wird wahrscheinlich eher als /a/ vom Hörer mis-klassifiziert (da es näher am Mittelpunkt von /a/ als am Mittelpunkt von /o/ ist) und daher nicht als Mitglied/Bestandteil von der Verteilung von /o/ betrachtet.

Auf S. 20 von samuels ist dasselbe Ausreißer-/o/-Token, aber diesmal sind die /o, a/ Verteilungen voneinander entfernter. Hier wird /o/ als Mitglied der /o/-Verteilung betrachtet (weil es näher an dessen Zentroid ist). Daher wird von /o/ der Mittelpunkt/ Zentroid neu berechnet, um diesen Ausreißer einzuschließen, und insofern verschiebt sich der Mittelpunkt von /o/ ein klein wenig in die Richtung von /a/ (d.h. der Mittelpunkt von /o/ mit Ausreißer ist näher an /a/ als der Mittelpunkt von /o/ ohne Ausreißer). Wenn nun (i) /a/ von /o/ ziemlich weit weg ist + (ii) ein Ausreißer zwischen den /o, a/ Verteilungen produziert wird, dann wird der Mittelpunkt von /o/ und daher die gesamte Verteilung in die Richtung von /a/ gezogen. Dies ist nicht eine funktionale, sondern eher eine statistische Erklärung für eine Kettenverschiebung.

### Prinzipien und Patterns

.....

NB: Patterns gestalten sich aus einer Kombination von Prinzipien.

I: 'in chain shifts, long vowels rise'

II: 'in chain shifts, short vowels fall'

Ia: 'the nuclei of upgliding diphthongs fall' z.B. /ei/ -> /ai/

III: 'back vowels move to the front'

Mehrere Beispiele für Prinzip III siehe in <http://www.phonetik.uni-muenchen.de/~jmh/papers/jphon11.pdf>, gleich in Absatz 1. Vor allem /u/-Vorverlagerung in verschiedenen Dialekten von England.

#### Pattern 1

Eine Kombination von I und II

z.B. der Great Vowel Shift im Englischen

[https://www.uni-due.de/SHE/SHE\\_Change\\_Phonological.htm](https://www.uni-due.de/SHE/SHE_Change_Phonological.htm)

siehe auch durian.pdf in unserer Lese-Webseite, S. 6

#### Pattern 2 (Kombination von I, II, III)

z.B. der Northern Cities Shift:

[http://www.ling.upenn.edu/phono\\_atlas/ICSLP4.html](http://www.ling.upenn.edu/phono_atlas/ICSLP4.html)

<http://www.pbs.org/speak/ahead/change/changin/>

'the fronting of back vowels': (caught -> cot)

'the raising of front vowels': 'cat' wird höher

'front short vowels fall' ('bit', 'bet' werden phonetisch offener).

#### Pattern 3

Kombination aus PI + PIII. Das beste Beispiel ist aus dem Schwedischen - siehe Labov (1994) S. 131 oben.

#### watt00.lvc.pdf

+++++

Kritik gegen Chain-shifting-Theorie. Man kann nicht alle Vokalverschiebungen allein durch intrinsische Faktoren erklären, sondern externe Faktoren wie 'social evaluation' (oben S. 70) spielen auch eine Rolle. Dann S. 70 unten: 'any attempt to explain sound change must take both speaker-centered (external) factors and system-

centered (internal) factors into account. A socially realistic model of vowel system dynamics must therefore allow for a balance to be struck between the autonomy of speakers on the one hand and that of the system on the other, since the relationship between speaker and system is better viewed as symbiotic than as antagonistic.'

Vielleicht dann die Details 71-3 überfliegen oder eher auf Tabelle 2 S. 74 + Text konzentrieren. Im Allgemeinen: Type I Monophthonge findet man auch in Schottland, Type III ist typisch für die Standardsprache in London/Südost England; Type II ist seltener und eher regional in Nordengland (Yorkshire, Derbyshire).

#### Das Experiment und Ergebnisse

Es ist eine auditive (nicht akustische!) Analyse der Spontansprache von 32 tyneside-englisch Sprechern, aufgeteilt in 2 Altersgruppen (S. 77) und 2 sozialen Schichten (working class/lower middle class). Es gab auch zwei Sprechstile: entweder eine gelesene Wortliste (WL) oder free-conversation (FC), S. 77. Die Daten und vor allem die FACE und GOAT Vokale wurden transkribiert.

Dann gleich zu den Hauptergebnissen (S. 86) springen, die anhand von den 3 Typen in Tabelle 2 erläutert werden.

Type I (Monophthonge) sind am häufigsten.

Type II nur in männlichen Working Class Sprechern, und sie verschwinden.

Type III nehmen zu, vor allem unter den lower-middle-classes (dies ist also eine Anpassung an Standardenglisch).

'Women appear to be leading the changes, and they are most advanced among middle-class speakers' - genau also wie Labov sagt. Punkte 6-7 auf S. 86 eventuell vergessen.

Es geht dann ('Face and goat as partner vowels') um Argumente für internal factors. Wenn dies ein internal shift ist, dann würden wir vielleicht eine Symmetrie erwarten, sodass die Änderungen zu FACE und GOAT gleichzeitig passieren (beide sind ja in diesem Dialekt auf derselben phonetischen Höhe). Dann vielleicht zu 'internal factors' auf S. 90 springen. Inwiefern also liegt Symmetrie vor (S. 91-92)? Vielleicht zu dem wichtigen Punkt dann auf S 92: 'The results of this study might therefore be taken as good evidence of internally motivated shifting in the TE vowel system. The principle of symmetry is the crucial factor conditioning the path of the change: it serves to keep changes in both vowels in line with one another.'

Aber dann werden die Probleme mit diesem Argument aufgelistet (um gegen internal factors zu argumentieren):

a. wenn der Lautwandel type I->II->III ist, dann müsste type I ziemlich selten sein, wenn type III erscheint. Aber type I ist sehr häufig.

Wenn type II sich aus type I entwickelt hat, ist es schwierig, die Asymmetrie zwischen Sprechern zu erklären (häufig in älteren Männern, selten in Frauen). Also Frauen (unten S. 92) scheinen von Type I zu Type III zu springen.

[Für mich sind das alles triftige Argumenten gegen die aufeinanderfolgende Entwicklung von type I>II>III, aber nicht zwingend gegen internal factors].

d.h. S. 93 'what is missing here is some account of the social marking that is attached to each of the variants'...'Rather, we need to take into consideration the probability that the source of the Type I and III variants is not internal but external.'

Dann Beweise/Argumente für externe Einflüsse unter anderem:

Type II nimmt ab unter middle-class Sprechern. Type II ist 'vernacular', also typisch für Tyneside-Englisch. Diese sind eventuell stigmatisiert. Also ist der Wandel soziologisch motiviert.

Type I nimmt zu wegen Kontakt mit nördlichen Dialekten.

Dann S. 95: dieses [ø:] nimmt zu, weil - ähnlich wie in Labovs Marthas Vineyard Daten - dies typisch für Tyneside-Englisch, nicht veraltet, und eventuell eine Reaktion gegen die Nivellierung von Tyneside-Englisch ist (diese Argumente finde ich sehr schwach). Vielleicht kommt [ø:] wegen Prinzip III zustande (internal factor), aber dann müsste man auch in Tyneside-Englisch eine Vorverlagerung von /u/ sehen, und das ist nicht der Fall.

Ihre Schlussfolgerung (S. 97): Es ist alles soziologisch bedingt... 'In conclusion, the changes described in this article may be understood as speaker-motivated adoptions and rejections of sociolinguistically marked surface forms brought about by differentials in social psychological attitudes...No appeal is necessary to spontaneous creation by the internal workings of the vowel system'

Neuseeland-Englisch (maclagan07.lvc.pdf)

+++++

Für Beispiele von NZE chain-shifts gibt es auch die Folie in vdata/Seminare/Prosody/lit/nze.pptx. Die vorderen kurzen Vokale sind höher geworden; sodass 'had' (START/TRAP) sich wie 'head' (DRESS) anhört, 'head' (DRESS) wie 'hid' (KIT), und 'hid' ist zentralisiert. Zusätzlich ist 'head' so hoch geworden, dass es mit 'heed' (FLEECE) überlappt. 'heed' hat jetzt einen deutlicheren On glide (mit späterem Target), ähnlich wie in Australisch-Englisch. Eventuell handelt es sich hier um einen push-chain, sodass DRESS aufgrund von TRAP-Raising weiter nach oben geschoben worden ist (S. 2); und sodass KIT wegen DRESS-raising in die Mitte geschoben worden ist. Die Frage ist, ob DRESS-raising und die Entwicklung vom On glide in FLEECE kausal zusammenhängen.

Es gibt übrigens einen Merger in NZE zwischen /eə, ɪə/ - 'bears' ist zu 'beers' geworden usw. d.h. /eə/ wurde mit DRESS nach oben gezogen; aber /ɪə/ ist nicht mit KIT zentralisiert geworden.

Siehe bitte an dieser Stelle die Kommentare zu NZE von torgerson.pdf (Pattern 4) auf S. 6 und 7 in der Lese-Webseite.

Die wichtigsten Ergebnisse:

1. (S8) 'for the younger, male nonprofessional speakers and the younger female speakers, dress and fleece become closer, until for the younger, nonprofessional female speakers, dress and fleece are in fact in the same position, and their ellipses overlap totally. For these groups, then, these two vowels occupy the same acoustic space.'

2. Fig. 8 zeigt die Diphthongisierung von FLEECE - d.h. die Vokalzielposition kommt spät im Vokal vor (der F2-Gipfel findet nicht zum zeitlichen Mittelpunkt statt, sondern später).

Dann würde ich die Ergebnisse überspringen bis 'Length' (S. 14).

Hier ist die Frage ob der Vokal in DRESS kürzer geworden ist. Eigentlich müsste das der Fall sein, aufgrund der intrinsischen Vokalhöhe (je phonetisch höher der Vokal, umso kürzer die Dauer nach Lehiste usw., also  $i < e < a$  in der Dauer usw.). Das Hauptergebnis: 'Because we have no reason to believe that 'strut' is changing, this suggests that voiceless dress may be shortening as it raises.'

Auf der anderen Seite haben Catherine Watson und ich gefunden, dass DRESS nicht mit zunehmender phonetischer Höhe kürzer wird: 'However, Watson et al. (2000) noted that as kit and dress rose for the four speakers they studied, they did not actually decrease in length.' Also man weiß eigentlich noch nicht so ganz genau ob DRESS-Raising mit dessen Kürzung verbunden ist. (Zusätzliche Möglichkeit für MA-Arbeit!).

Dann 'voiced and voiceless phonemes'

Die Autoren führen recht teleologische Argumente durch, also 'Fleece seems to be reacting to the encroachment of dress on its space by becoming more diphthongal'.

S. 18 'If the diphthongization of fleece is causally related to the height of dress, we would predict that the diphthongization should be most advanced in the voiceless tokens'

Weil sich im stimmlosen Kontext FLEECE und DRESS am meisten überlappen, müsste FLEECE im stimmlosen Kontext am meisten diphthongisiert sein. Das Ergebnis: 'Figure 14 shows box plots of the degree of diphthongization of voiced (left panel) and voiceless (right panel) tokens, for both older and younger speakers. It shows that there is more diphthongization in the voiceless tokens than the voiced tokens, and that the voiceless tokens are increasing in diphthongization more rapidly than the voiced ones.'

Die Diskussion: hier ist die Frage, ob Längenunterschiede überhaupt noch im NZE zum Vorschein kommen.

Die wichtigste Schlussfolgerung (S21/22: 'However, in New Zealand English, length is not sufficient to distinguish fleece (and, particularly voiceless fleece) from dress. It is threatened in F1-F2 space, and also threatened in the length dimension. It seems reasonable to assume that vowels most likely to react in a chain-shift are those that are threatened in all three dimensions. Thus, voiceless fleece diphthongizes most vigorously, because all three dimensions are under pressure...') (S23) 'The diphthongization of fleece is most advanced in tokens that are followed by voiceless codas. These are the tokens that are most endangered by the high dress, as they are then distinguished neither in acoustic space, nor by length.'

## # 20. The perception of phonetic contrasts by infants

### Allgemeine Methodik

+++++

head turning paradigm

<http://ling.uni-konstanz.de/pages/home/grijzenhout/bsl/index.php?cont=method>  
sowie

[https://www.youtube.com/watch?v=EFlixifDk\\_o](https://www.youtube.com/watch?v=EFlixifDk_o)

### werker2002

+++++

Dies ist ein Reprint von Werker & Tess (1984). Eine Folie dazu in /vdata/Seminare/Prosody/lit/werkertess.pptx

Hier wird der Befund behandelt, dass Kleinkinder im Alter von 6-8 Monaten (KK-6) den Unterschied zwischen phonetischen Lauten erkennen können, auch wenn diese Unterschiede in ihrer eigenen Muttersprache nicht vorkommen. Aber diese Fähigkeit geht allmählich verloren: Erwachsene und Kinder im Alter von 3-4 Jahren sind nicht mehr so fähig, solche Unterschiede wahrzunehmen. Eventuell hat das etwas mit der Überlagerung des Phonem-Systems der Muttersprache zu tun - oder vielleicht mit dem Zeitpunkt, zu dem Kinder die Beziehungen zwischen Wörtern und Objekten lernen (wie in der nächsten Studie von Stager & Werker, 1997 gezeigt wird).

Das wichtigste aus der Einleitung:

'In our previous work (Werker et al., 1981), we compared English adults, English infants, and Hindi adults on their ability to discriminate two pairs of Hindi speech contrasts that are not used in English. One of the Hindi contrasts was the dental voiceless aspirated vs. the voiced aspirated voicing distinction /tha/-/dha/ and the other, the voiceless, unaspirated retroflex vs. dental place of articulation distinction /ʈa/-/ta/. The results showed that infants aged 6–8 months can discriminate these sounds as well as Hindi adults, but that English adults cannot, particularly in the case of the place of articulation contrast. Furthermore, English children at ages 12, 8, and 4 were as poor as the English adults in terms of their discriminative performance (Werker & Tees, 1983), showing the decline to be evident in children as young as 4 years of age.'

### Experiment 1

NB was sie  $\tilde{k}$ ,  $\tilde{q}$  transkribieren = k', q' also velare und uvulare *Ejektive* (mit glottalem statt pulmonalem Luftstrommechanismus).

Inwiefern können KK-6 den Salish-Unterschied [k'a, q'a] und das Englische [ba, pa] erkennen?

Methode: Head-turning paradigm (siehe oben)

Ergebnisse: Abb. 2

### Experiment 2 und 3

Hier wollten die Autoren genauer feststellen, über welchem Zeitraum die Fähigkeit den Unterschied wahrzunehmen, verloren geht. Diesmal wieder die obigen Kontraste sowie Hindi /ta, ʈa/ getestet.

Das Ergebnis:

## Abbildung 4

## Diskussion

'In summary, these experiments provide strong support for the claim that young infants can discriminate many of the phonetic distinctions used across natural languages without relevant experience, and that there is a decline in this ability as a function of specific language experience.'

stager97.nature.pdf

+++++

siehe auch Folien 5-16 von Lecture5-Sounds of Words /vdata/Seminare/Prosody/lit/Lecture5-SoundsOfWords.pptx. Das sind Folien von <http://www.socsci.uci.edu/~lpearl/>

Hier wird versucht, die Frage zu beantworten, weshalb diese phonetische Diskriminierungsfähigkeit zwischen 6-8 und 10-12 Monaten verloren geht. Die allgemeine Antwort: zu viele phonetische Details können störend sein, wenn Kinder anfangen, Wörter mit externen Objekten zu verknüpfen.

## Die Methodik

KK (Alter 14 Monate) lernen Wort-Objekt-Beziehungen (*Habituation*)

Test: bemerken sie einen Wechsel im Objekt, im Wort, oder in beiden?

Diese Untersuchung.

(i) Wir wissen, dass 8-monatige Kinder zwischen phonetischen Lauten differenzieren können, auch wenn diese nicht in der Muttersprache vorkommen.

(ii) Ebenfalls wissen wir, dass diese Fähigkeit allmählich verloren geht.

(iii) In dem Alter, zu dem diese Fähigkeit verloren geht, lernen Kinder Objekte mit Wörtern zu assoziieren.

Die Hauptfrage:

Sind (ii) und (iii) irgendwie miteinander verbunden?

Experiment 1: Habituation. 2 Wort-Objekte-Paare (bih, dih). D.h. Es gibt ein bih-Objekt und ein dih-Objekt. In einem Test ('same Trial') hören KK das richtige Wort: sie sehen bih-Objekt und hören [bi] = bih-Objekt-[bi]. Ebenfalls dann dih-Objekt-[di].

In einem zweiten Test hören sie das falsche Wort. D.h., sie sehen das (gelernte) Objekt für bih und hören [di]. Dies ist bih-Objekt-[di]. Oder dih-Objekt-[bi].

Es wird gewechselt von bih-Objekt-[bi] auf bih-Objekt-[di]  
oder von dih-Objekt-[di] auf dih-Objekt-[bi]

Ist dieser Wechsel den KK aufgefallen? Nein!

## Experiment 2

Das gleiche wie Expt. 1, aber Kinder lernen nur bih-Objekt. Ihnen wird bih-Objekt-[bi] und dih-Objekt-[bi] präsentiert. Merken sie den Unterschied, wenn von bih-Objekt-[bi] nach bih-Objekt-[di] gewechselt wird? Die Antwort: KK im Alter von 14 Monaten nicht, aber KK im Alter von 8 Monaten schon. Die Erklärung. KK-14, aber nicht KK-8 lernen Wort-Objekt-Assoziierungen und durch dieses Lernen werden feine phonetische

Unterschiede nicht wahrgenommen. KK-8 nehmen diese Unterschiede wahr, weil sie Wort-Objekt-Beziehungen noch nicht gelernt haben.

### Experiment 3

Das gleiche wie Expt. 2, aber diesmal ist der phonetische Unterschied zwischen den Wortpaaren viel ausgeprägter: also sie lernen ein lif-Objekt und die Aufgabe ist, ob sie den Unterschied zwischen lif-Objekt[lif] und lif-Objekt[nim] bemerken. Hier müsste KK-14 diesen Unterschied wahrnehmen. Das war auch so. Dies bestätigt die Schlussfolgerung im Expt. 2: kleine phonetische Unterschiede werden nicht wahrgenommen, wenn Kinder die Beziehung zwischen Wörtern und Objekten lernen.

### Experiment 4

Könnte es aber sein, dass (a) KK-14 einfach verlernt haben, phonetische Details wahrzunehmen? Oder (b) ist es eher der Fall, dass sie phonetische Details nicht so gut lernen können, wenn sie Wort-Objekt-Beziehungen lernen müssen?

Hier haben KK bih-Schachbrettmuster gelernt. In der Testphase hörten sie dann bih-Schachbrettmuster und dih-Schachbrettmuster. Hier gibt es also kein Erlernen von Wort-Objekt-Beziehungen. Und hier waren KK-14 genauso fähig wie KK-8 zwischen /b, d/ zu diskriminieren.

D.h., das Problem für KK-14 in Expt. 2 ist tatsächlich, dass durch das Erlernen der Beziehung zwischen Wörtern und Objekten feine phonetische Details (wie der Unterschied zwischen /b, d/) ausgeblendet werden können. 'This finding eliminates the possibility that infants of 14 months can no longer make fine phonetic discriminations. Instead, it presents compelling evidence that it is only when infants are attempting to learn the meaning of words that they fail to attend to the fine phonetic information.'

## #21. Perceptual magnet model

kuhl95.pdf

+++++

Die Erfahrung mit der Sprache verändert, wie die Sprachperzeption verarbeitet wird. Die Beziehung zu dem Vortrag Thema 20: Kleinkinder im Alter von 6 Monaten (KK-6) können generell unabhängig von der Sprache zwischen Sprachlauten differenzieren; aber ab ca. KK-14 (Kleinkinder im Alter von 14 Monaten) geht diese Fähigkeit verloren, sodass Kinder kaum mehr in der Lage sind, zwischen Sprachlauten zu differenzieren, die in ihrer Sprache nicht vorkommen.

Bei dem Perceptual-Magnet-Modell (PMM) führt die Erfahrung mit einer Sprache zu einer perzeptiven Verzerrung, und zwar auf eine solche Weise, dass es schwieriger wird, zwischen zwei prototypischen (P) im Vgl. zu nicht-prototypischen (NP) Lauten zu differenzieren.

P für /i/: das wäre etwas ähnliches wie der Mittelwert aller wahrgenommenen /i/-Vokale.

NP: das wäre dann eine Produktion, die ziemlich am Rande der /i/-Verteilung ist. Eine gute Abbildung dazu ist Fig. 13.3 in hawkins.pdf

Fig. 2: Die synthetischen Stimuli für P und NP-Vokale: also die Stimuli in denselben Abständen von den P- und NP-Mittelwerten.

Sowohl Erwachsene als auch KK-6 zeigen einen PM-Effekt (Fig. 3). Dies zeigt, wieviele Tokens als äquivalent mit dem P oder NP-Vokal beurteilt wurden (deutlich höher für P).

Dann ist die Frage, inwiefern dieser Magnet-Effekt mit Erfahrung verbunden ist. Dazu dann der Test mit Amerikanisch-Englisch und Schwedisch (die Stimuli dazu in Abb. 4). Die Ergebnisse (Abb 5) zeigten einen sprachspezifischen Magnet-Effekt. Die wichtigste Schlussfolgerung: sogar bei KK-6 verursacht die Erfahrung mit der Sprache einen Magnet-Effekt.

Motherese

.....

Vor allem Eigenschaften von Vokalen werden übertrieben: Motherese überträgt also laut dieser Theorie übertriebene Merkmale eines *Prototypen* - was ja für das Kind den L1-Spracherwerb erleichtern soll. Daher (S. 131): 'motherese ... may "tutor" infants on the sound patterns of their native language.'

Erwachsene und PM-Effekt

.....

Expt. 1

Versuchspersonen mussten für ein Kontinuum zwischen 'he' und 'hay' pro Stimulus einschätzen, wie prototypisch es für /i/ war.

Expt 2

Diskriminierungsexperiment. Die Hörer konnten nicht so gut zwischen zwei P- vs. zwei NP-Laute differenzieren (Fig. 8).

### Expt 3

Multidimensional scaling (MDS): dies ist eine statistische Analyse, um perzeptive Entscheidung graphisch umzuwandeln, sodass zwei Laute umso näherliegender in der entstehenden Graphik sind, je ähnlicher sie einander perzeptiv sind.

Die Vpn. hörten alle 13 Stimuli paarweise mit unterschiedlichen Dauern (ISI) zwischen den Stimuli und mussten sagen, ob die Paare gleich oder unterschiedlich waren. Das graphische MDS-Ergebnis davon in Fig. 9: die Tokens (ganz oben) hatten denselben akustischen Abstand; aber /i/ und teilweise /ei/ (wie hay) Tokens näher am Prototyp waren schlechter differenzierbar.

### NLM-Modell

.....

Kuhl geht also davon aus, dass Kinder mit denselben allgemeinen auditiven Fähigkeiten geboren sind, um ein akustisches Kontinuum aufzuteilen (Fig. 10A, S. 140): 'the NLM theory holds that what is "given by nature" is the ability to partition the sound stream into gross categories separated by natural boundaries' (S. 139/140).

Und:

'it is important to note... that the boundary effects shown in Fig. 10A are not entirely unique to humans; some are also displayed by nonhuman animals. The theory Kuhl is developing thus claims that the infant's ability to hear the relevant differences between phonetic units is innate and attributable to general auditory processing mechanisms.'

Kuhl widerspricht hiermit einem der wichtigsten Schlagworte der Motor Theory of Speech Perception, nämlich der Behauptung 'speech is special'. Es ist laut Kuhl eben *nicht* eine sprachliche Fähigkeit, mit der Kinder geboren werden, um ein Kontinuum aufzuteilen, sondern diese Fähigkeit entsteht aufgrund von auditiven Mechanismen (des Hörsystems, das alle gesunden Kinder gemeinsam haben). Daher .. S. 141 'Given that this acoustic space is initially divided by natural psychophysical boundaries... '

Was gelernt wird ist also der PM-Effekt, wodurch diese allgemeine auditive Aufteilung aufgrund der Erfahrung mit der Muttersprache verzerrt wird (Fig. 10 B, S. 140).

### Natural language magnets and the perception of foreign language sounds

.....

Nachdem die Wirkung des perceptual magnets zunimmt, sind notwendigerweise einige Laute in anderen Sprachen nicht mehr so unterscheidbar. Das ist so, weil der perceptual magnet den auditiven Raum in Richtung der Vokalverteilung der Muttersprache verzerrt. Damit will Kuhl auch die Ergebnisse von Janet Werker erklären (dass KK-14 es schwieriger finden, zwischen Lauten zu diskriminieren, die nicht in der Muttersprache vorkommen). Ebenso jene Ergebnisse (S. 142 unten), dass Erwachsene, die eine Fremdsprache lernen, Schwierigkeiten haben, einen L2-Laut genau richtig zu produzieren, wenn der L2-Laut sich in unmittelbarer Nähe zu einem L1-Laut befindet (z.B. das deutsche und englische /i/ sind ähnlich, aber nicht genau identisch, vgl. 'lief', und 'leaf'. Aber L1-Englisch-Sprecher oder L1-Deutsch-Sprecher haben Schwierigkeiten,

das /i/ der anderen Sprache genau zu produzieren, gerade weil das englische und deutsche /i/ so ähnlich - also so nah am jeweiligen Prototypen - sind).

Alles ab S. 144 ist eventuell nicht ganz so wichtig, und kann weggelassen werden.

kuhl08

+++++

Die Bilder sind hier:

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2606791/figure/fig1/>

Ich würde mit diesem anfangen:

4. 'NATIVE LANGUAGE MAGNET THEORY, EXPANDED' (S. 5)

Die Grundprinzipien des NLM

(i) 'Distributional patterns and infant-directed speech are agents of change'

Im Allgemeinen entsteht ein Magnet aufgrund der Erfahrung in der frühen Kindheit mit der Sprache. Vor allem sind zwei Sorten von Erfahrung wichtig:

(a) die Häufigkeit ähnlicher Sprachlaute.

Wenn Sprachlaute häufig in einem gewissen Teil des auditiven Raumes wahrgenommen werden, dann wird der Raum dementsprechend verzerrt. Dies ist eigentlich nichts Neues. Z.B. wird /i/ überwiegend mit einem hohen F2 und niedrigen F1 produziert, sodass das Kind viele Laute mit diesen Eigenschaften hört. Deswegen entsteht ein Magnet, also eine perzeptive Verzerrung in diesem Bereich.

(b) Motherese (child-directed speech) übertreibt die (proto-)typischen Eigenschaften - daher diskriminieren Kinder Sprachlaute effektiver, wenn sie viel Motherese hören (viele Evidenzen dafür gibt es allerdings nicht).

(ii) 'Language exposure produces neural commitment that affects future learning'

Dafür gibt es auch nicht gerade sehr viele Beweise. Aber es soll wieder der Fall sein, dass es einen Vorgang gibt, wobei 'initial language exposure causes physical changes in neural tissue and circuitry that reflect the statistical and perceptual properties of language input.'

(iii) 'Social interaction influences early language learning at the phonetic level'

Dies ist interessanter (und deutet darauf hin, dass Fernsehen Lautwandel nicht verursachen kann).

3 Gruppen von L1-englischen Kleinkindern (a, b, c)

(a) KK, die Mandarin-Chinesisch von L1-Mandarin-Chinesisch Sprechern hörten

(b) KK, die nur Englisch hörten

(c) wie (a), aber sie hörten Mandarin-Chinesisch entweder vom Fernsehen oder vom Tonband.

In der Testphase, in der Kleinkinder Mandarin-Chinesische Silben identifizieren mussten, schnitt (a) besser ab als (b, c). d.h. Spracherwerb kann nicht passiv, z.B. vom Fernsehen, erlernt werden: die (soziale) Interaktion ist wichtig.

(iv) The perception-production link is forged developmentally

Dazu hat Kuhl eigentlich nicht sehr viel zu sagen, da sie kaum mit Sprachproduktion (nur Perzeption) arbeitet. Im Allgemeinen aber:

Die Perzeption steuert die Produktion teilweise durch Imitation:

'listening to simple vowels in the laboratory alters infants' vocalizations, and that this ability emerges at approximately 20 weeks of age but is not present at 12 or 16 weeks of age.'

(v) Die frühe Perzeption beeinflusst die spätere Sprachentwicklung.

Das hatten wir schon teilweise mit Motherese - also laut dieser Theorie können Kinder, die in ihrer Kindheit viel Motherese hatten, später Sprachlaute besser diskriminieren.

Das ERP-Experiment

.....

Keineswegs auf die Details eingehen. Die Methode ist in /vdata/Seminare/Prosody/lit/method21.pdf

Das wichtigste Ergebnis:

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2606791/figure/fig4/>

oder /vdata/Seminare/Prosody/lit/topic21fig4.pdf

Kleinkinder, die effektiver zwischen L1-Kontrasten diskriminieren, lernen später die Sprache effektiver, d.h. bei ihnen vergrößert sich später der Wortschatz schneller. Umgekehrt: Kleinkinder, die effektiver zwischen L2-Kontrasten diskriminieren (also besser in der Lage sind, zwischen Lauten zu differenzieren, die in ihrer Sprache nicht vorkommen), lernen die eigene Sprache nicht so effektiv (Wortschatz wächst langsamer). (Schlechte Nachrichten also für ein phonetisches Wunderkind!).

Daher (unter Discussion): 'Infants' performance on both the native and non-native phonetic discrimination tasks, measured using ERPs at 7.5 months of age, significantly predict children's language abilities 2 years later, but differentially. Better native phonetic abilities predict faster advancement in language, whereas better non-native phonetic abilities predict slower linguistic advancement...Rather, infants' abilities to learn phonetically from exposure to language predict the rate of language growth over the first 30 months of age.'

-----

Das wichtigste aus dem Rest des Artikels (ab S. 15):

Die 5 Phasen (aber das ist eigentlich nur eine Wiederholung mit ein paar zusätzlichen Details)

Phase 1: 'infants discriminate all phonetic units in the world's languages.'

Phase 2: die Häufigkeit der Verteilungen von Sprachlauten im auditiven Raum + Motherese führen zu diesen perceptiven Verzerrungen in Richtung der Lautverteilung der Muttersprache. In dieser Phase ist die soziale Interaktion wichtig; und es entstehen Perzeptions-Produktions-Beziehungen.

Phase 3: (Hier hat Kuhl selber kaum etwas beigetragen). Das ist die Phase, in der Phonotaktik, Prosodie, und Wortschatz-Zuwachs stattfinden.

Phase 4: Nur die Verzerrungen, die aufgrund von L1 stattgefunden haben, sind stabil, daher 'exposure to a new language does not automatically create new neural structure'.

-----

## 5. Vorhersagen

a. Bilingualismus: Phase 2 bei bilingualen Kindern könnte länger dauern

b. soziale Interaktion führt zu robusteren Sprachkenntnissen

c. 'critical learning hypothesis.' Diese (von Lenneberg stammende) Hypothese besagt, dass sich die neuronalen Funktionen ab einem gewissen Alter festsetzen, sodass Zweitspracherwerb ab diesem Alter viel schwieriger wird (weil die neuronale Plastizität verloren geht). Das wichtigste hier: 'at the cusp of learning, phonetic perception of native and non-native contrasts is negatively correlated. This supports the idea that learning itself may play a role in reducing the future capacity to learn new phonetic patterns. D.h. laut Kuhl beeinträchtigt der Spracherwerb der Muttersprache zunehmend die Fähigkeit, die Laute einer zweiten Sprache zu erlernen.

## # 22 Speech learning model

Die wichtigsten Sache aus flege95.pdf

+++++

The critical period hypothesis (CPH)

.....

Diese Theorie ist angewandt worden, um zu erklären, weshalb der fremdsprachliche Akzent bei Erwachsenen im Vgl. zu Kindern ausgeprägter ist. CPH geht zum großen Teil auf Lennebergs (1967) Theorie im 'biological foundations of language' zurück, dass die benötigte neuronale Plastizität für den automatischen Erwerb einer Sprache ab ca. dem 12 Lebensalter eingeschränkt wird. Und daher müsste es ab diesem kritischen Alter nicht mehr möglich sein, einen L2 ohne fremdsprachlichen Akzent zu erlernen. Gegen diese Hypothese argumentiert Flege vehement. Hier ist Fig. 1 aus flege95.pdf sehr relevant: wenn der CPH greift, müsste die Wahrnehmung des fremdsprachlichen Akzents nach diesem kritischen Alter stark zunehmen (man müsste also einen Knick in der Verteilung sehen). Statt dessen zeigt die Graphik, dass der fremdsprachliche Akzent abnimmt, je länger man Zeit in der L2-Umgebung verbringt.

SLM (speech learning model)

+++++

Der SLM befasst sich mit der Beziehung zwischen den L1- und L2-Lauten auch in bilingualen Sprechern. Dieses Modell wird am deutlichsten in Tabelle I in flege95.pdf dargestellt.

Vielleicht H1 und H2 überfliegen (oder kurz erwähnen), dann näher eingehen auf H3 und H4

H1: die Beziehung d.h. gegenseitiger Einfluss von L1- und L2-Sprachlauten ist phonetisch und nicht phonologisch. z.B. (siehe H5) verschiebt sich Aspiration in L1-Deutsch durch das Erlernen von L2-Französisch - und Aspiration ist in deutsch nicht phonologisch (nicht kontrastiv). L1-japanische Sprecher (S. 240) können Englisch /l, r/ voneinander meistens in initialer Position nicht trennen. Final geht das, weil finaler /l/ velarisiert ist, und ein velarisierter /l/ und /r/ sind akustisch sehr unterschiedlich. Dies bestätigt wieder, dass die L1-L2 Einflüsse basieren auf \*phonetische\* und nicht auf grobe \*phonologische\* Unterschiede.

H2: Damit sich in der L2-Sprache ein neue phonetische Kategorie bildet, muss der Laut von anderen phonetischen Kategorien in der Muttersprache perzeptiv differenzierbar sein. Z.B. wird sich wahrscheinlich eine neue phonetische Kategorie für nasalisiertes [ɛ̃] (fr. 'main', Hand) bilden, da es in deutsch keine ähnliche Kategorien gibt. Aber deutsche haben Schwierigkeiten das englische [a] ('bat' = Fledermaus) zu lernen, weil es phonetisch zu nah an dem deutschen [ɛ] (Bett) ist.

H3. Phonetische Entfernung zwischen L1- und L2-Lauten

.....

Je phonetisch entfernter ein L2-Laut von irgendeinem L1-Laut, umso wahrscheinlicher ist es, dass dieser L2-Laut gelernt wird (= dass eine getrennte phonetische Kategorie für

L2-Laut etabliert wird). z.B. Deutsch und Englisch haben einen sehr ähnlichen aber nicht identischen /i/ (leaf, lief). Daher fällt es deutschen oder englischen Muttersprachlern sehr schwer, diesen Laut in der anderen Sprache akzentfrei zu lernen (sie haben eher daher die selbe phonetische Kategorie für beide Laute). Laut dieser Hypothese müsste es einfacher sein, für deutsche Muttersprachler das englische KV-5 in 'bath' zu lernen, da CV-5 in den meisten deutschen Dialekten nicht vorkommt.

#### H4. AOL (age of learning)

.....

Dies ist H4 in flege95.pdf. AOL betrifft das Alter, mit dem man erstmal L2 lernt. Je niedriger, umso wahrscheinlicher ist es, dass man phonetische Unterschiede (wie in H3 dargestellt) zwischen ähnlichen L1- und L2-Lauten aufdecken kann. (Daher haben Kinder einen geringeren fremdsprachlichen Akzent im Vgl. zu Erwachsenen).

#### H5 und H6. Gegenseitige Störung von L1- und L2-Lauten

.....

Hier ist die Hypothese, dass die L1 und L2-Räume kognitiv nicht getrennt sind, was gegenseitige Einflüsse zwischen den Sprachen zur Folge haben kann. Daher könnte sogar L1 von L2 betroffen werden. Daher 'Flege (1987a) found that bilinguals produced stops in their L1 with VOT values resembling those typical for stops in the L2'. Das würde z.B. bedeuten: wenn L1-deutsche Muttersprachler L2-französisch lernen, wird VOT in L1-deutsch kleiner (da VOT von /p t k/ kürzer ist in Französisch im Vgl. zu deutsch). [Das Beispiel, das Flege von Mack, 1990 anführt, ist mir nicht ganz verständlich].

Die Haupthypothese bei H6 ist, dass die L1- und L2-Laute im selben kognitiven Raum gespeichert werden. Zweitens müssen phonologische Kategorien in diesem Raum voneinander Abstand halten, und dies kann zur Folge haben, dass deswegen ein L2- (oder L1) Laut verschoben wird (im Vgl. zu den entsprechenden Lauten bei Muttersprachlern). z.B. ist /u/ in Englisch sehr frontiert und daher nah am deutschen /y/. Man könnte sich daher laut dieser Theorie vorstellen, dass ein bilingualer deutsch-englischer Sprecher /y/ etwas frontierter produziert, und/oder /u/ etwas rückverlagerter, damit sich das deutsche-/y/ und das englische-/u/ nicht komplett überlappen.

H7 ist nur: man produziert die Laute, wie sie im kognitiven Raum gespeichert sind. Das ist keine Hypothese - vielleicht weglassen.

Vielleicht könnte man dann 1-2 mehr Beispiele für H3-H6 (oder H3-H7) in den nächsten 3 Abschnitten finden (production and perception of vowels; production and perception of initial consonants; production and perception of final consonants), die ich auf gar keinem Fall ausführlich zusammenfassen würde. Der letzte Teil wirft auch viel mehr Fragen als Antworten auf, und kann überflogen/weggelassen werden.

Die wichtigsten Sache aus flegejob1997.pdf

+++++

flege95.pdf befasst sich hauptsächlich mit den gegenseitigen L1-L2-Einflüssen. Das wesentliche hier und im nächsten Paper, ist weshalb Kinder eher als Erwachsene in der Lage sind, den fremdsprachlichen Akzent im L2-Erwerb zu verlieren.

Vielleicht also hier CPH (critical period hypothesis) sowie die Daten in Fig. 1 von flege95.pdf erwähnen - um zu zeigen, dass CPH dieses Phänomen - dass Kinder den fremdsprachlichen Akzent verlieren - nicht erklären kann (S. 171 von flegejop1997.pdf)

Statt dessen wird geprüft, ob L1-Verwendung den fremdsprachlichen Akzent in L2 beeinflusst (unter der Hypothese: je mehr L1 verwendet wird, umso größer ist bei L2 der fremdsprachliche Akzent).

Hier vielleicht auf das 'single system hypothesis' von Flege näher eingehen (was ja mit H5 und H6 in flege95.pdf verbunden ist):

d.h. The single system hypothesis ... predicts that the loss of the L1, or its attenuation through disuse. Put another way, the "less" L1 there is, the smaller will be its influence on the L2 (S. 171/172).

...

According to the single system hypothesis, bilinguals are unable to fully isolate the L1 and L2 phonetic systems, which necessarily interact with one another..

according to the single system hypothesis, the phonic elements of the L1 subsystem necessarily influence phonic elements in the L2 system, and vice versa. (S. 172)

Für die Untersuchung gab es zwei Gruppen HiUse (eine Gruppe, die viel L1-italienisch spricht) vs. LoUse (relativ wenig italienisch wird gesprochen). Die Hypothese: der fremdsprachliche Akzent müsste für HiUse ausgeprägter sein.

Zusätzlich (und dies ist weniger interessant/relevant) wird geprüft, ob L1-Sprecher aus Ottawa-Kanada besser in der Lage sind, einen fremdsprachlichen Akzent im Vgl. zu L1-Sprecher aus Alabama-USA aufzudecken. Diese Hypothese beruht auf die Idee, dass L1-Sprecher aus Ottawa Kanada eher in der Lage sein müssten, dies zu tun, da sie in einer zweisprachigen (Engl-Franz.) Kultur leben.

Hier (sehr) kurz auf die Methode eingehen. Vor allem betonen, dass 'The subjects in the NI groups were matched according to their age of immigration to Canada from Italy, but differed according to self-reported use of Italian'

Dann sehr schnell die Ergebnisse überfliegen. Eventuell reicht es, Fig. 1 S. 178 zu zeigen. Die danach folgende Sache mit A und A' scores würde ich vielleicht weglassen. Die wichtigste Schlussfolgerung: mehr fremdsprachlicher Akzent bei der Gruppe HiUse. Eigentlich minimale (tendenzielle) Unterschiede zwischen den zwei (L1-Ottawa und L1-Alabama) Hörergruppen.

Die wichtigsten Sache aus flege06.jop.pdf

+++++

dies ist eine ähnliche Thematik wie flegejop1997.pdf. Hier wird geprüft, ob length-of-residency (LOR) in der L2-Umgebung die Stärke des fremdsprachlichen Akzents beeinflusst. Das neue hier: die Vpn. hatten unterschiedliche Age-of-arrival (AOA) in der fremdsprachlichen Umgebung - und ähnliche LOR (z.B. einen Vergleich zwischen zwei Personen im Alter von 10 und 20 Jahren, die beide 3 Jahre in der fremdsprachlichen Umgebung wohnten). Und es gab zwei Gruppen mit unterschiedlichen LORs von 3

Jahren oder 5 Jahren. Also das wichtigste hier: AOA und LOR werden zum ersten Mal in dieser Studie voneinander entkoppelt.

Hier wieder die Methode überfliegen, und die Ergebnisse eventuell anhand von Fig. 1 zeigen. Die wichtigsten Ergebnisse: LOR hat keinen Einfluss aber AOA schon (Kinder haben einen geringeren fremdsprachlichen Akzent im Vgl. zu Erwachsenen mit dem selben LOR). Ebenfalls wird gezeigt, dass der Aufenthaltsdauer keinen Einfluss auf den fremdsprachlichen Akzent hat (die Ergebnisse von Time 1 und Time 2 sind fast gleich).

Dann auch zu diesem wichtigen Ergebnis auf S 167 springen:

"For both NK adults and children, significant moderate correlations existed between sentence ratings and overall percentage use of English (the more the use of English, the milder the foreign accents)" ... und "We speculate that age-related differences in English use may have contributed to the difference in strength of foreign accent observed for the NK children and adults." (S. 170).

d.h. wie in flege1997.jop.pdf: je mehr man die L2-Sprache verwendet (unabhängig davon ob man Kind oder Erwachsen ist), umso geringer der fremdsprachliche Akzent.

Flege gibt aber nicht eine präzise Erklärung für den Grund des geringeren fremdsprachlichen Akzents bei Kindern im Vgl. zu Erwachsenen:

"If child immigrants hear the L2 more than adult immigrants and actively use the L2 more, this might result in better L2 pronunciation, especially if the (heard) L2 input comes from L2 native speakers. This hypothesis will need to be investigated carefully in future research" S. 170).

## # 23 Speaking style and variation

schillingestes.pdf

+++++

Aus diesem Paper würde ich auf folgendes konzentrieren:

- Section 1: What is style. (Die Definition stilistischer Variation).

Vor allem ist Stil die Anpassung des Sprechstils an das Hörerpublikum (an die Person, mit der man spricht): z.B. wird die Aussprache etwas bairischer, wenn man mit Marktleuten in einem bayerischen Dorf spricht.

- Section 3. 'Attention to speech'. Vor allem das Modell von Labov und 'attention to speech' (Section 3, S. 378/379). Hier darauf fokussieren: es gibt in Labovs Modell eine Beziehung (1) zwischen der Formalität der Sprache (casual->formal) und inwiefern der Sprecher seine eigene Sprache beobachtet (der eigenen Sprache Aufmerksamkeit schenkt). (2) zwischen sozialer Klasse und stilistischer Variation (Abb. 15.1, 15.2). d.h. (a) je mehr die Sprache stilistisch Richtung 'formal' neigt, umso mehr werden Varianten verwendet, die für eine obere Klassenschicht typisch sind; daher wie auch in Abb. 15.1 und 15.2 gezeigt wird, wird die niedrigste soziale Gruppe am meisten zwischen nicht-formalen und formalen Sprechsituationen ihre Aussprache stilistisch verändern. [Eventuell diesen Teil mit labov72.pdf unten zusammen behandeln]. Allerdings: das könnte wohl stimmen für amerikanisch-englisch, aber inwiefern das wirklich auf andere Dialekte/Sprachen übertragbar ist, weiß man nicht so ganz genau.

Ich würde dann 3.2 weglassen.

3.3 Eventuell kurz auf die Kritik von Bell (1984) eingehen - dies betrifft am meisten (1) oben, dass es sehr schwer ist, eine Beziehung zwischen Aufmerksamkeit und Formalität der Sprache zu quantifizieren. Mehr Details zu diesem wichtigen Kritikpunkt kann man aus bell84.pdf S. 147-150 in der Webseite entnehmen.

4. Audience design. Einige der wichtigsten Punkte zu audience design schnell/knapp zusammenfassen.

Audience design besagt, dass stilistische Änderungen sich nicht nach der Aufmerksamkeit der eigenen Sprache richten, sondern eher nach dem Publikum/Hörern mit dem man spricht (z.B. aus 4.1). d.h. (siehe auch bell84.pdf, S. 159) die Personen, mit denen man sich unterhält, haben einen viel größeren Einfluss auf stilistische Variationen als in Labovs Modell. Es gibt auch eine Beziehung zwischen Bells audience design model und Konvergenz im Speech Accommodation Theory - siehe bell1984.pdf S. 162 für mehr Details. [Vielleicht dann Bells Theorie zu audience design mit der Studie von hay.pdf unten verknüpfen].

4.2 bis zum Ende von schillingestes.pdf würde ich vielleicht weglassen.

labov72.pdf

+++++

vielleicht hier auf die sehr interessante und innovative Methode eingehen, um die Beziehung zwischen stilistischer Variation und sozialer Klasse festzustellen. Es handelt sich um Sprechdaten aus 3 Kaufhäusern in New York (Saks, Macys, Kleins) die mit einer

abnehmenden sozialen Schicht verbunden sind. Gegenstand der Untersuchung ist ein /r/, der eher vor einer niedrigen sozialen Schicht vokalisiert wird. Zwei Varianten von /r/ wurden untersucht entweder vor einem Konsonanten (fourth) oder final (floor). Der Sprechstil wurde manipuliert, indem Labov mehrere Verkäuferinnen eine Frage stellte (die die Antwort 'fourth floor' produzierte) und dann so tat, als hätte er die Verkäuferin nicht verstanden (sodass die Verkäuferin 'fourth floor' in einem klareren Sprechstil wiederholen musste). Die Ergebnisse: viel mehr /r/-Tilgung in Kleins als in Macys als in Saks (Abb. 13.1)

Das interessante an Abb. 13.2 ist, dass in emphatischer/klarer Sprache Macys nah an Saks kommt - insofern wird eine Verbindung zwischen stilistischer Variation und sozialer Klasse hergestellt (je deutlicher gesprochen wird, umso wahrscheinlicher ist es, dass die Sprache in die Richtung einer gehobeneren sozialen Schicht tendiert). [es ist nicht klar aus Abb. 13.2, aber wenn ich mich nicht täusche, ist 4th floor links in 13.2 in einer normalen Sprache, und 4th floor rechts in einer emphatischen Sprache (= wiederholte Sprache der Verkäuferin). Dies wird auch irgendwie aus Tabelle 13.1 ersichtlich. z.B. wenn man 'd' und non-data ignoriert, hat Macys-Casual für floor  $48/(48+62) = 44\%$  von r-1 (= /r/ wird produziert) und Macys-Emphatic für floor ist  $31/(31+20) = 61\%$  also einen Anstieg von ca. 17% in der Verwendung von /r/ wenn 'floor' von Macys-Verkäuferinnen mit Emphase produziert wird - daher die Zahlen 44 un 61 in Abb. 13.2.

hay99.pdf

+++++

Diese Studie ist ein Test einiger Aspekte der Theorie von Bell (1984) zu audience design. Diese Theorie besagt nicht nur, dass der gewählte Stil eines Sprechers sich nach dem vorhandenen Publikum richtet, sondern auch dass der eigene Sprechstil von anderen Personen beeinflusst wird, die zwar nicht präsent sind, aber besprochen werden/im Dialog salient sind.

In der Studie wird die Sprecherin und auch das vorhandene Publikum konstant gehalten (wichtig da sie auch einen Einfluss auf Stil haben) durch die Wahl von Oprah Winfrey als chat-show host. Die zu untersuchende Variable war Monophthongisierung von /ai/, die für African-American-English typisch ist. Als Hypothese wurde getestet, dass Winfray zunehmend von dieser Monophthongisierung bei der Vorstellung von afrikanisch-amerikanisch (AA) Sprechern Gebrauch machen würde. Dies wäre konsistent mit Bells Theorie (nicht anwesende Sprecher können den Stil beeinflussen, wenn sie besprochen - oder wie hier - vorgestellt werden). Dies wurde auch gefunden ('an AA referee significantly increases the probability of monophthongisation (pdf S. 2 oben rechts).

Zusätzlich wurde die Verbindung zwischen lexikalischer Häufigkeit und Monophthongisierung untersucht - und es konnte festgestellt werden, dass /ai/ in häufigeren als in selteneren Wörtern eher monophthongisiert wurde. Dies ist konsistent mit vielen Studien, die in dem Paper erwähnt werden, dass Lautwandel zuerst in häufig vorkommenden Wörtern stattfindet, eventuell weil sich in häufig vorkommenden Wörtern stilistische Variationen am meisten durchsetzen: 'Conclusion': 'Our results... support a model where it is highly frequent words which emerge as the best candidates for the display of speaker style.' d.h. die Anpassung des Sprechers an das Publikum in Bells audience-design model wird eventuell eher in häufigeren Wörtern zum Vorschein kommen.

redford09.pdf

+++++

Es geht hier darum, wie stilistische Variation von Kindern erworben werden. Hier wird die Hypothese geprüft, dass Kleinkinder im Alter von ca. 2 Jahren nicht fähig sind, die Klarheit des Sprechstils an die Sprechsituation anzupassen (z.B. verdeutlichen sie nicht nach einer Klärungsfrage was sie gesagt haben).

Viele Details sollen überflogen werden. Hier das wichtigste.

Es handelt sich um 30 amerikanisch-englische Schuldkinder, jeweils 10 Kinder im Alter von 3, 4, und 5 Jahren. 30 Studenten beurteilten, wie klar die Kinder gesprochen hatten.

Es gab zwei Sprechsituationen:

- i. Clear speech, in dem Kinder Bilder benannt haben, und in dem die Kinder aufgefordert wurden, mit maximaler Klarheit zu sprechen.
- ii. Casual speech: in dem Kinder + Eltern zusammen mit Holzblöcken spielten, auf denen die Bilder geklebt wurden, die in clear speech gesprochen worden waren. Im Spiel ermutigten die Eltern zwischendurch die Kinder, die auf den Holzblöcken geklebten Bilder zu benennen.

Die Studenten (3 Gruppen von 10) beurteilten die clear- und casual- gesprochenen Wörter der Kinder für Wortklarheit, Klarheit des Vokales, und wie erwachsen die von den Kindern produzierten Wörtern waren.

Die Vokale der Kinder wurden auch für verschiedene Merkmale akustisch analysiert (hier vielleicht nicht zu sehr auf die Details auf S. 1439 eingehen).

Die Hauptegebnisse der Hörerurteile (Listener Ratings)

.....

Hier wirklich nur auf das wichtigste eingehen also:

Fig.1 Man sieht deutlich einen Trend:

- i. Vokalklarheit, Wortklarheit, und wie erwachsen das Wort war, nehmen zu von 3 bis 4 bis 5 Jahren.
- ii. 3 jährige Kinder - im Gegensatz zu 4 und 5 jährigen - hörten sich genau so kindisch an in clear und casual speech: daher überlappen die beiden grauen Punkte (clear, casual) unten links komplett - aber dies ist nicht der Fall für die 4 und 5 Jährigen. Den gesamten Rest - inkl. Fig. 2 - würde ich dann weglassen.

Dann zur Akustik. Was ist hier wichtig? Wie folgt

1. Dauer

Abbildungen 3 und 4 würde ich auch vergessen - das ist zu komplex und uninteressant.

Das wichtigste:

- i. 3-year-olds produced longer vowels than did 4- and 5-year-olds
- ii. 3-year-olds also showed the most variability in vowel duration.

2. Pitch

Hier eventuell gar nichts erwähnen.

### 3. Vokale.

Hier nicht Fig. 6 präsentieren sondern gleich zur Schlussfolgerung: Stil hatte keinen Einfluss auf die Formanten.

### 4. Konsonanten

Eigentlich sind die Ergebnisse genau so uninteressant wie in 3..

i. Plosive werden in klarer Sprache eher gelöst (Tab. 2) aber es gibt keine Interaktion mit Alter (= der Unterschied zwischen clear und casual war nicht größer bei 5 als bei 3 Jährigen).

ii. Wenn klarer gesprochen wurde, verschiebt sich das Spektrum von /s, ʃ/ (Sibilanten) nach oben in clear vs. casual speech (eventuell weil Luft schneller durch die Verengung getrieben wird) wie Abb. 7 zeigt. Zusätzlich gehen die Werte clear und casual am meisten auseinander bei 5 Jährigen (bei 3 Jährigen gehen die Mittelwerte zwar auseinander, aber die Verteilung überlappen sich stark). Die zeigt daher, dass eventuell nur 5-jährige Kinder imstande waren, die clear-casual Speech Unterschiede via Sibilantenunterschiede zu übertragen.

### Diskussion und Ergebnisse (S. 1445):

.....

Hörer konnten nicht zwischen klaren und informellen Äußerungen dreijähriger Kinder unterscheiden; aber dann doch bei 4 und 5 Jährigen. Diese Hörerergebnisse waren mit den Produktionsergebnissen konsistent (4 und 5 jedoch nicht 3 jährige Kinder trennten zwischen 'clear' und 'casual' Sprechstilen zumindest teilweise in ihrer Produktion). Dann wird (S. 1446) darauf eingegangen, inwiefern sich die stilistischen Änderungen in Kindern von Erwachsenen unterscheiden. z.B. komprimieren Erwachsene viel mehr als Kinder den Vokalraum in 'casual speech'. Ebenfalls produzieren Erwachsene eine große  $f_0$ -Auslenkung in klarer Sprache (weil das Wort stark akzentuiert wird); das ist nicht der Fall bei Kindern. Sie senken statt dessen ihr  $f_0$  wahrscheinlich weil die Aufforderung klar zu sprechen als 'ich muss wie ein Erwachsener reden' interpretiert wird (und Erwachsene haben einen tieferen  $f_0$ , daher senken sie  $f_0$ ).

## # 24: Phonotactics and language acquisition

cutler92.jml.pdf

+++++

Hier wird die Theorie begründet, dass Hörer die rhythmischen Eigenschaften von Wörtern verwenden, um Wortgrenzen aufzudecken. Insbesondere gibt es hier einige Beweise, dass Hörer vor einer sogenannten starken Silbe - also einer Silbe mit primärer oder sekundärer lexikalischen Betonung - eine Wortgrenze hypothetisieren: d.h. sie beginnen das mentale Lexikon vor jeder starken Silbe zu durchsuchen. (Die gilt übrigens nur für germanische Sprachen oder Sprachen mit einem trochäischen (stark-schwach) Wortbetonungsmuster).

Konsistent damit: die meisten mehrsilbigen Wörter in englisch haben eine stark-schwache rhythmische Struktur also eine trochäische und keine jambische (weak-strong) Struktur.

Die neuen experimentellen Untersuchungen hier.

1. Eine Untersuchung von 'slips of the ear', in denen Wortgrenzen hinzugefügt werden (wenn man etwas falsches wahrnimmt). Hier wurde bestätigt, dass es viel wahrscheinlicher ist, dass eine Wortgrenze vor einer starken (analogy -> and allergy) als vor einer schwachen Silbe (effective -> effect of). hinzugefügt wird. Ebenfalls gab es konsistent mit der Hypothese eine größere Wahrscheinlichkeit, dass Wortgrenzen vor schwachen (my gorge is -> my gorgeous) als vor starken (is he really -> Israeli) Silben getilgt werden

[Vielleicht dann Expt 2 überfliegen oder weglassen aus Zeitgründen].

cutler1994.lingua.pdf

+++++

Hier wird jetzt die obige Idee auf die Wahrnehmung von Wortgrenzen von Kleinkindern erweitert. Ebenfalls werden Sprachen mit unterschiedlichen rhythmischen Strukturen berücksichtigt. Was hier neu ist:

Rhythmus in anderen Sprachen (S. 92, 93)

.....

In Französisch soll die Silbe für die Aufdeckung von Wortgrenzen viel wichtiger als in Deutsch sein. Wenn z.B. französische Hörer die Silbe /bal/ identifizieren müssen, dann tun sie das viel schneller in 'balcon' als in 'balance', wahrscheinlich weil in 'balance' (aber nicht in 'balcon') der /l/ zur zweiten Silbe gehört. Und sie reagieren schneller auf /ba/ in 'balance' als in 'balcon' weil in balance jedoch nicht in balcon die Silbengrenze gleich nach /ba/ vorkommt.

(Das komplizierte Beispiel von japanisch ist interessant aber nicht sehr gut durchdacht/erklärt. Ich würde das weglassen).

wichtig, S. 93: 'cross-linguistic differences in speech segmentation are characteristics of the listeners and do not simply follow from the nature of the speech input.' d.h. Sprecher setzen die rhythmische Wortsegmentierungsstrategie ein, die sie in der Muttersprache gelernt haben. = Franz. Sprecher, die Englische Sätze lesen, verwenden eher eine Wortsegmentierungsstrategie, die auf Silben basiert ist.

Bilinguale Sprecher verwenden die Wortsegmentierungsstrategie ihrer dominanten Sprache (S. 94).

Bezüglich Kinder

.....

S. 96 befasst sich mit dem Erwerb von Sprachrhythmus. Es wird vermutet, 'the rhythmic segmentation procedures used by adult listeners exist not for purposes of adult processing at all, but are simply traces which remain from a period when the segmentation problem dominated the infant's language processing' d.h. diese Fähigkeit Sprachrhythmus für die Aufdeckung von Wortgrenzen einzusetzen, ist ein Nebenprodukt des Erlernens des Sprachrhythmus in der frühen Kindheit. Scheinbar lernen bilinguale Kinder nicht zwei rhythmische Strategien in den jeweiligen Sprachen sondern nur eine in der dominanten Sprache.

S. 97 'What is remarkable about this process is that it seems to happen only once, if the evidence from the bilingual studies is reliable. That is, exposure to two differing rhythmic regularities (syllabic and stress rhythm, for instance) does not result in the ability to use both types of rhythm in speech segmentation; a language user appears to be able to command only one rhythmic segmentation procedure.'

jusczyk1993childdev.pdf

+++++

Drei sehr wichtige Experimente:

Expt 1: L1-englische-KK-9 hören länger auf trochäische als auf jambische Wörter zu.

Expt 2: Dies war nicht der Fall für KK-6. Scheinbar muss diese rhythmische Fähigkeit in dem Zeitraum ab 6 bis 9 Monate gelernt werden.

Expt 3: Das gleiche Ergebnis wie bei Expt. 1 aber mit tiefgepassten Stimuli, d.h. in denen die phonetischen Einheiten zerstört wurden. KK-9 abstrahieren daher die rhythmischen Verhältnisse unabhängig von dem segmentellen Material - oder anders herum: der Erwerb der segmentellen Phonetik (also von Konsonanten und Vokalen) ist nicht eine Voraussetzung für Sprachrhythmus-Erkennung (siehe auch houston2000psychbullrev.pdf).

mattyscogsci1999.pdf

+++++

Das Ziel hier ist die Bedeutung phonotaktischer Reihenfolgen sowie Wortrhythmus für die Aufdeckung von Wortgrenzen von KK-9 zu testen.

Zwei Sorten von phonotaktischen Beschränkungen wurden untersucht:

(a) Between-Reihenfolgen, die häufig an Grenzen aber kaum wortintern vorkommen = between word. Ein Beispiel wäre /mg/, was in deutsch nur an Morphem/Wortgrenzen erscheinen kann (Stammgast).

(b) Within-Reihenfolgen, die viel häufiger wortintern als an Grenzen vorkommen. Ihr Beispiel für english ist /ŋk/. Ein zusätzliches Beispiel für deutsch wäre vielleicht die reduzierte Form /bm/ in 'geben' = /gebm/, da es in Deutsch keine Wörter (wegen Auslautverhärtung) mit finalem /b/ geben kann.

.....

Die Haupthypothese: KK-9 bevorzugt ein sogenanntes integriertes Wortform d.h. bei zweisilbigen Wörtern (i) mit einem trochäischen Betonungsmuster (aus all den bereits oben diskutierten Gründen) und (ii) mit Phonemreihenfolgen, die wortintern typisch sind.

.....

#### Experiment I

KK-9 hörten CVC.CVC in dem C.C entweder einen (a) between- oder (b) within-Cluster überbrückte. Die Wortbetonung war trochäisch (also Hauptbetonung auf der ersten Silbe). Die Hypothese: KK-9 hört für eine kürzere Zeit auf (a) zu (= hört (a) nicht so gerne), weil die wortinterne phonologische Reihenfolge für englische Wörter atypisch ist. Dies war auch der Fall (Fig. 1 S. 475).

#### Experiment II

das gleiche aber mit einem jambischen Wortbetonungsmuster. Die Hypothese hier ist etwas kompliziert und wie folgt:

Zweisilbige jambische Wörter haben einen weak-strong Betonungsmuster. Der Beginn von strong ist ein Zeichen für eine Wortgrenze (aus all den Gründen, die Cutler erläutert). Daher hört man eher eine Wortgrenze # in einer Reihenfolge von einer schwachen + starken Silbe also weak-#-strong. Diese Wahrscheinlichkeit eine Wortgrenze zu hören, wird bei between-clusters nochmal gestärkt, da z.B. /mg/ nicht wortintern sein kann.

1. Daher weak-#-strong mit C.C als 'between' gibt eindeutige Cues für eine Wortgrenze vor 'strong'

2. Dagegen ist weak-strong mit C.C als within zweideutig. Weil:

- weak-strong vermittelt den Eindruck: es gibt hier eine Wortgrenze.
- within gibt den Eindruck: es gibt keine Wortgrenze.

Weil weak-strong mit C.C als within inkonsistent bezüglich Wortgrenzen sind, müssten KKs sie ablehnen oder eher weak-#-strong mit C.C als between bevorzugen. Das war auch der Fall (Fig. 2).

#### Experiment III

Die selben stimuli als bei Expt. I aber eine Pause wurde zwischen den Silben hinzugefügt. Dadurch müssten KK-9 deutlich eine Wortgrenze hören. d.h. strong-Pause-Weak wird als strong-#-Weak wahrgenommen (mit Wortgrenze) Dann sind die weiteren Argumente wie bei Expt II, also: within gibt den Eindruck: es gibt keine Wortgrenze

Weil strong-Pause-weak mit C.C als within inkonsistent mit der Wahrnehmung von Wortgrenzen sind, müssten KKs sie ablehnen, oder eher strong-Pause-weak mit C.C als between bevorzugen. Das war auch der Fall (Fig. 3).

#### Experiment IV

[kann weggelassen werden - ich finde das alles nicht sehr überzeugend]

Hier soll geprüft werden, ob Wortrhythmus oder phonotaktische Beschränkungen für die Aufdeckung von Wortgrenzen wichtiger ist.

Die Stimuli

(a) trochäisch mit between-word clusters.

trochäisch vermittelt die Wahrnehmung eines einzigen Wortes, between deutet jedoch auf eine Wortgrenze und daher 2 Wörter

(b) jambisch mit within-word clusters

jambisch vermittelt den Eindruck von 2 Wörtern mit einer Wortgrenze, within jedoch auf keine Wortgrenze.

Obwohl (a, b) beide zweideutig sind (ob eine Wortgrenze vorkommt oder nicht), bevorzugten KK-9 (a). Dies soll zeigen, dass die interne Konsistenz von Wörtern eher durch Wortrhythmus als durch die Phonotaktik verletzt wird.

mattys2001cognition.pdf

+++++

Hier wird auf eine direktere Weise geprüft, inwiefern KK-9 seltene wort-interne phonologische Reihenfolgen für die Erkennung von Wortgrenzen verwenden.

Die Experimente im Allgemeinen. CVC Logatome (gaff, tove) wurden in Sprechmaterialien eingebettet. CVC wurde von C umringt also C.CVC.C. Es gab zwei Sorten von Materialien.

(a) = present = mit guten phonotaktischen Beschränkungen. Hier sind C.C eher 'between' im Sinne vom vorigen Paper - sie kommen also nicht so oft/kaum wortintern vor. Das CVC Wort müsste also besonders salient sein, da es durch die beiden 'between' C.C isoliert wird.

(b) = absent = mit schlechten phonotaktischen Beschränkungen. Hier kommt C.C eher wortintern vor. Daher ist das CVC-Logatom schwieriger aus dem Materialien perceptiv zu segmentieren (weil die phonotaktische Information nicht ausreicht, um das CVC-Logatom zu isolieren).

Experiment I

In der Testphase hörten KK-9 verschiedene Logatome, darunter diejenigen die in den Trainings-Materialien vorkamen. Es war ein head-turning procedure. Die Hypothese: KK-9 müsste länger auf die Logatome schauen, nachdem sie (a) im Vgl. zu (b) gehört haben - da bei (a) die Merkmale, um das Wort zu identifizieren, salienter waren. Dies war auch der Fall (die linken zwei Balken in Abb. 1).

Experiment II

Genau das gleiche wie Expt. I mit einem Unterschied.

In dem (a) present Bedingung wird hier der CVC nur initial durch phonotaktische Beschränkung von einem davor kommenden C getrennt. d.h. es war C1.CVC.C2 und C1.C war between, jedoch C.C2 war within. Der Hypothese bei diesem Test: Wortgrenzen werden aufgedeckt hauptsächlich aufgrund der Trennung von Wort-onsets (d.h. der Beginn von einem Wort ist wichtiger als das Ende von einem Wort, um das Wort aus einer kontinuierlichen Äußerung zu segmentieren - dies ist übrigens konsistent mit einigen Theorien von 'lexical access' wie Marslen-Wilsons cohort Modell).

Das wichtigste Ergebnis(Fig. 2)

"Overall, the patterns of results were identical to those obtained in Experiment 1...These results demonstrate that a spoken stimulus can be efficiently segmented ... when only its onset is phonotactically cued."

Experiment III

(dieses Experiment ist etwas schwach und könnte weggelassen werden)

das gleiche aber diesmal wurde der Logatom-Offset anstatt Onset durch phonotaktische Beschränkungen (also durch einen Between-Cluster) getrennt.

Das Ergebnis: ähnlich wie für Expt II. d.h. KK-9 ist besser in der Lage, das Wort zu erkennen, nachdem es Sprachmaterialien mit C.C-Offset Beschränkungen wahrgenommen hat. Jedoch wird das Wort nicht ganz so gut erkannt wie in Experimenten I und II:

"However, in certain respects, the results of the present experiment are weaker than those of the previous two experiments."

## #25. Prosodic boundaries and language acquisition

christopheDevSci2003.pdf

+++++

Es geht hier teilweise um die Beziehung zwischen Syntax und Prosodie. Jedoch ist der Hintergrund nicht so eindeutig erklärt!

Wir fangen vielleicht mit Prosodie an - und lassen die Syntax weg.

Prosodie

.....

Im französischen wird die rechte Grenze von einer Prosodie gestärkt. Nehmen wir an wir produzieren den Satz wie auf S. 214 als zwei prosodische Phrasen [ ]

[le grand orang-outang ] [ était enervé]

dann sind die Silben 'tang' und vé von enervé sehr prominent/deutlich, weil sie unmittelbar vor der rechten prosodischen Phrasengrenze ] auftreten.

In Turkish ist das anders herum. Hier wird nicht die rechte sondern die linke Grenze von einer prosodischen Phrase gestärkt.

daher wenn man eine Phrasengrenze:

[yeni kitabimi] [almak istiyor]

nach kitabimi setzt, dann sind yeni und almak prominent - weil sie gleich nach einer [ linken Phrasengrenze vorkommen. Allerdings sind es die letzten Silben dieser phraseninitialen Wörter, die prominent sind, weil die Wortbetonung im türkischen final ist.

Das wichtigste also: französisch macht die rechte Grenze von einer prosodischen Grenze stark, türkisch die linke.

Übrigens ist der Satz im Türkischen:

yeni = neu

kitabimi = Buch

almak = kaufen

istiyor = sie will

Die Frage ist, ob Kleinkinder (KK) die Unterschiede zwischen diesen unterschiedlichen prosodischen Mustern im Franz. und Türkischen wahrnehmen können.

Um das zu testen, wurde 'reiterant speech' verwendet, um den phonologischen Inhalt zu entfernen. Reiterant speech: alle Phoneme in beiden Sprachen wurden mit /j, l, m, e/ ersetzt, und zwar auf eine solche Weise, dass sie phonologisch sehr ähnlich sind. Die Sätze wurden auch resynthetisiert, um alle anderen akustischen Parameter gleich zu halten.

Fig. 2 ist nur ob Erwachsene zwischen diesen Phrasen unterscheiden können. Offenbar ja.

Zu den Kindern. High amplitude sucking paradigm. Ihnen wurden verschiedene Reihenfolgen der türkischen und französischen (reiterant, resynthetisierten) Phrasen vorgespielt. Wenn sie einen Unterschied dazwischen hören, dann wird mehr gesaugt usw. Hier ist das Ergebnis in Fig. 3. Die Reihenfolge unterschiedlicher Sprachen (dunkelgrau) führt zu einer Änderung im sucking rate, die Reihenfolge derselben Sprachen (hellgrau) nicht.

Daher hören Kleinkinder die prosodischen Unterschiede zwischen den französischen und türkischen Sätzen.

Jetzt zur Syntax

.....

Oft haben Sprachen mit einer Stärkung der linken prosodischen Grenze eine Reihenfolge Objekt-Verb (Complement-Head) also nicht wie im englischen Verb-Objekt (Head-Complement) wie 'She saw Susan' (Subject-Verb-Obj) sondern im türkischen 'She Susan saw' (Subj-Obj-Verb)

Daher die Hypothese: Kinder nehmen diese prosodischen Unterschiede wahr (ob links-stark oder rechts-stark) weil das über die typische Reihenfolge von head-complement in der Sprache aussagekräftig sein könnte (also: wenn sie in einer Sprache aufwachsen, in der sie ständig eine Verstärkung der linken prosodischen Grenze wahrnehmen, dann ziehen sie eher die Schlussfolgerung, die Sprache hat die Reihenfolge Objekt-Verb also Complement-Head).

Übrigens 'head' in Syntax ist der obligatorische Teil einer syntaktischen Phrase daher ist Verb der head von Verb-Phrase (weil ohne Verb kann es keine Verb-Phrase geben), Nomen der head von Noun-Phrase usw..

christophe-Infancy2001.pdf

+++++

Dieses Paper ist klarer. Der Hintergrund: sind Kinder imstande eine prosodische Grenze wahrzunehmen? Dies wurde vorhin anhand von französischen Materialien getestet, und zwar ob Kinder den Unterschied zwischen [mati] und [ma][ti] erkennen, wo ][ eine prosodische Phrasengrenze ist. (Dies wurde durch Materialien getestet, in denen /mati/ im selben Wort vorkam (mathematicien) oder nicht (cinema titanisque = cinema] [titanisque). Die phonetischen Merkmale von ][ (= von einer prosodischen Phrasengrenze) sind nicht unbedingt eine Pause, sondern eine Längung der Silbe vor ] (=phrasenfinale Längung) und eine Längung des ersten Konsonanten nach [ (= wortinitiale Stärkung) der Phrasengrenze.

Sogar Kindern im Alter von 3 \*Tagen\* konnten diesen Unterschied erkennen.

In diesem Experiment wird geprüft, ob so etwas auch in einer anderen Sprache wahrnehmbar ist. d.h. ob französische Kleinkinder den Unterschied im Spanischen zwischen [lati] und [la][ti] wahrnehmen.

Table I kann man wirklich überfliegen. Es ist nur eine Bestätigung, dass es akustisch-phonetische Merkmale gibt, die [lati] von [la][ti] trennen. Das wichtigste Ergebnis ist das

Bild rechts in Fig. 1 (das Bild links in Fig. 1 würde ich eventuell gar nicht präsentieren) und der Text "This experiment showed that French newborns discriminate between Spanish bisyllabic stimuli that differ in whether or not they contain a phonological phrase boundary".

Control in Fig. 1 rechts ist wenn die Kinder eine Reihenfolge unterschiedlicher Stimuli aber mit der selben prosodischen Grenze bekamen z.B. [lati] gefolgt von [sana].  
 Experimental group: sie bekommen eine Reihenfolge von Stimuli mit unterschiedlichen prosodischen Grenzen wie [lati] gefolgt von [la][ti]. Offenbar ist der Amplitude von sucking höher im Fall der experimentellen Gruppe. Daher erkennen Kinder den Unterschied zwischen [lati] und [la][ti]

juszyk92cogpsych.pdf

+++++

Prosodische Phrasengrenzen kommen oft (aber keineswegs immer!) an Stellen vor, an denen es größere syntaktische Grenzen gibt. z.B. zwischen einer Nomenphrase und Verbphrase '[Der kluge Mann ] [hat die Prüfung bestanden]. ][ ist eine prosodische Phrasengrenze. Und dies ist konsistent mit Syntax da NP = Der kluge Mann, VP = hat die Prüfung bestanden.

Vielleicht können Kinder daher das Auftreten prosodischer Phrasengrenzen verwenden, um Syntax zu parsen d.h. um einen Satz in größere syntaktische Einheiten aufzuteilen.

Wenn KK darauf eingestellt sind, dann müssten sie prosodische Phrasengrenzen bevorzugen, wenn sie an Stellen vorkommen, wo es größere syntaktische Brüche gibt, wie an der Grenze zwischen einer NP und VP. Sie lehnen aber eher Sätze ab, wenn eine prosodische Phrasengrenze mitten in einer syntaktischen Einheit vorkommt. Das ist was hier getestet wird, indem Pausen an gewissen Stellen in Sätzen hinzugefügt werden - z.B. entweder an NP-VP-Grenzen, oder innerhalb von NP- oder VP-Phrasen. Der Test ist head-turning also wie lange Kleinkinder den Kopf in die Richtung des wahrgenommenen Stimulus drehen.

#### Experiment 1

.....

Kinder im Alter von ca. 9 Monaten (KK-9) hörten zwei verschiedene Sorten von Texten  
 Coincident version: hier wird eine Pause an NP-VP Grenzen hinzugefügt. Daher eine Pause zwischen 'you' = NP und 'spill your cereal' = VP.

Noncoincident version. Pausen wurden mitten im VP hinzugefügt. Daher nach 'spill' in der Verbphrase 'spill your cereal'

Ergebnisse: infants listen longer to strings of English that are segmented at the major phrase boundary, just prior to the main verb of the predicate, as opposed to strings that are segmented within the predicate phrase itself.

#### Experiment 2

.....

das gleiche Experiment mit KK-6.

Ergebnis: for the 6-month-old infants, the difference in orientation times between the Coincident and Noncoincident versions was not significant

### Experiment 3

.....

das gleiche wie Expt. I aber bei non-coincident wird diesmal die Pause mitten in einer Nomenphrase (NP) statt mitten in einer VP hinzugefügt... Siehe Tabelle 3 auf pdf S. 17 (S. 268). Es gibt also zusätzliche Materialien wie;

The dangerous ][ wild animals

also eine Phrasengrenze mitten im NP 'the dangerous wild animals'

Hauptergebnis: es spielt keine Rolle ob die Pause im NP oder im VP vorkommt (beide sind gleich schlecht). Daher 'the pauses in the Noncoincident versions occurred within the subject-noun phrase or within the predicate-verb phrase indicated no significant differences between these two locations'

### Experiment 4

.....

Hier wird geprüft, ob diese Unterschiede zwischen Coincident und Non-coincident speech genauso vorkommen in child-directed als in adult-directed speech.

Child-directed speech = die Sprache, die verwendet wird, wenn man mit einem Kleinkind spricht.

Die Hypothese: eventuell kommt der Unterschied zwischen coincident/non-coincident deutlicher in child-directed speech zum Vorschein, da die Prosodie ausgeprägter ist.

Hauptergebnis: Although infants in both the Child-directed and Adult-directed conditions displayed significant preferences for the Coincident versions, the effects appeared to be stronger for the Child-directed samples.

### Experiment 5

+++++

Vielleicht achten Kleinkinder auf prosodische Phrasengrenzen viel mehr, gerade weil sie sie dazu verwenden, um zu lernen, wie die Syntax strukturiert ist. Erwachsene brauchen dagegen prosodische Phrasengrenzen kaum (mehr), um syntaktische Brüche aufzudecken, weil sie ihre Kenntnisse der Sprache dazu einsetzen können. Wenn das so ist, dann reagieren vielleicht Erwachsene nicht so sehr auf den Unterschied zwischen 'coincident' und 'non-coincident': d.h. sie finden beide solche Sätze gleich natürlich in der Produktionen der Sprache. Die Materialien waren ähnlich wie in Expt. 4 aber die Sätze wurden tief-pass gefiltert, sodass die Wörter nicht erkennbar waren.

Das Ergebnis (contra die Hypothese): The present results indicate that adults judged the Coincident versions of the samples to be "more natural" than the Noncoincident versions. This suggests that the division of the utterances into parts at the major phrasal boundary between subject and predicate constitutes a more natural perceptual partitioning of the utterance for them than the alternative positions within the verb or subject noun phrases. Thus, at least for the present set of materials, adults are attentive to acoustic correlates of major phrasal units.

Der Rest von diesem Section kann eventuell weggelassen werden: es ist nur eine wiederholte Darstellung der möglichen akustischen Cues der prosodischen Phrasengrenzen. Vielleicht gleich zum Ergebnis springen, S. 281. 'In summary, the preliminary sorts of acoustic analyses that we were able to conduct on our materials do

suggest the presence of acoustic markers of phrasal units. In particular, there are clear pitch drops prior to phrase boundary breaks for the Coincident versions of both the Storybook and Spontaneous speech samples. In addition, for the Storybook materials, there is some indication that the average duration of the syllable immediately prior to a pause is longer in the Coincident versions than in the Noncoincident versions of the samples. Thus, we now have evidence that there are potential acoustic markers of phrasal units in our materials, and that infants listen longer to samples that are consistent with these markers.'

#### Experiment 6

Hier wird geprüft, ob der Unterschied in der Präferenz für coincident speech wirklich wegen akustisch-prosodischer Information zustande kommt. Wenn ja, dann müssten KK-9 genauso auf die Unterschiede zwischen coincident und non-coincident speech reagieren, auch wenn die Sprachmaterialien wie im Expt. 5 tiefpass-gefiltert wurden, sodass keine phonetischen Informationen übrigbleiben.

Hauptergebnis: "Thus, the present results show that even with low-pass filtering, the preference for the Coincident versions of the samples still remains.

#### Experiment 7

kann weggelassen werden. Hier wird geprüft, ob KK-6 vielleicht doch auf coincident vs. non-coincident mit anderen (mit storybook) Materialien reagieren (die Antwort: nein).

## # 26. First language acquisition of sociophonetic information

Ich würde im wesentlichen auf 2 Papers konzentrieren: foulkes05.language.pdf (und die anderen Papers von Foulkes nur als Zusatzmaterial dazu verwenden); und roberts97.childlang.pdf. Sehr viel Literatur wird von Foulkes in den Papers besprochen; das kann zum großen Teil übersprungen werden. Auch die Exemplartheorie in foulkes06.jop.pdf kann überflogen werden.

foulkes05.languagepdf

+++++

Es geht hier um:

1. child-directed speech (also die Sprache, die verwendet wird, wenn Erwachsene mit Kindern sprechen)
2. ob in child directed speech (CDS) mehr Standard- im Vergleich zu Dialekt (vernacular) Varianten verwendet wird
3. inwiefern die /t/-Produktion in Kindern von Erwachsenen abweicht
4. inwiefern 2. variiert, wenn man mit Mädchen vs. Jungen spricht.

Hier wird Newcastle-Englisch analysiert, teilweise weil die große phonetische Variation in der Produktion von /t/ mit sozialer Klasse verbunden ist.

Die Analyse

.....

Männliche und weibliche Kinder sowie ihre Mütter wurden aufgenommen im Alter zwischen 2- 4 Jahren. Alle Kinder stammten aus einer Arbeiterklasse.

/t/ in den Kindern sowie /t/ im CDS der Mütter wurden analysiert. CDS der Mütter wurde mit der /t/-Verwendung in Gesprächen von Erwachsenen verglichen (PVC Datenbank S. 185).

Die Analyse befasst sich insbesondere mit /t/ zwischen Sonoranten (water, winter, bottle), und finalem /t/ vor Vokalen (get in, hat on). S 185

Es wird hauptsächlich unterschieden, wann ein [t] - als ein stimmloser alveolarer Plosiv - produziert wurde, und wann glottalisiert wurde - also [ɖ] - d.h. ein stimmhafter /d/ mit Knarrstimme. Ein Beispiel von einem Spektrogramm von [ɖ] in Newcastle-Englisch ist im nächsten Paper foulkes06.jop.pdf S. 1, Fig. 1.

Die Produktion von [ɖ] ist für young working-class women typisch (S. 186)

/t/ zwischen Sonoranten

.....

Das wichtigste Ergebnis für /t/ zwischen Sonoranten Fig. 1 S. 187. Hier sieht man dass

1. im CDS viel mehr [t] (Standard) als [ɖ] im Vgl. zu Erwachsenengesprächen verwendet wird. Ein ähnliches Ergebnis wurde von Parsons (2000) Fig. 2 S. 188 gefunden.
2. in der CDS-Sprache verwenden Väter mehr [ɖ] als Frauen - also neigen Väter viel mehr dazu, Dialekt in CDS zu verwenden.

3. Fig. 3, S. 191: in der CDS-Sprache wird mehr [t] im Vgl. zu [d] verwendet, je älter das Kind.

#### Wortfinaler /t/

.....

Hier gibt es 4 phonetische Varianten

Standard [t] (stimmloser Plosiv) [wie oben]

[d] [wie oben]

[ɹ] ein /r/ ähnliches Laut (postalveolar Approximant)

[r] (auch als [ɹ] im Paper transkribiert: ein Alveolarer Tap)

#### Die wichtigsten Ergebnisse

Fig. 4: Viel mehr [t, d] als [ɹ, r] in CDS vs Erwachsenersprache ([ɹ, r] sind sehr stark mit Dialekt/vernacular verbunden)

Fig. 5: Mehr [t] wenn Mütter mit Töchtern reden im Vgl. zu Müttern mit Söhnen

Fig. 6: eine Abnahme in der Verwendung von [t] in CDS mit zunehmendem Alter des Kindes

#### Allgemeine Interpretation

.....

1. Es könnte sein, dass mehr [t] in CDS verwendet wird, weil [t] klarer/deutlicher ist im Vergleich zu den anderen Varianten d.h. vielleicht wollen Mütter die Sprache für die Kinder im CDS verdeutlichen. Jedoch lehnen die Autoren diese einfache Interpretation vielleicht mit Recht ab, weil eventuell das Erlernen des /t/s dadurch nicht unbedingt vereinfacht wird. Der Grund: Erwachsene - und dies betrifft vor allem /t/ in wort-medialer Position - verwenden kaum [t] sondern viel mehr andere Varianten also [d]: daher ist es schwer zu verstehen, wie durch häufiges Verwenden von [t] in CDS das Erlernen der Sprache vereinfacht wird.

2. Statt dessen wird vermutet, dass das häufige Verwenden von [t] aus soziophonetischen Gründen zustande kommt. d.h. Mütter wissen, dass [t] mit einer höheren sozialen Schicht/mit mehr Prestige assoziiert ist (Erwachsene aus der Arbeiterschicht verwenden überwiegend andere Varianten wie [d]), und Frauen wollen ihren Kindern diese Prestige-Variante beibringen. Daher taucht [t] häufig in CDS auf.

3. Eventuell konsistent mit einer soziophonetischen Erklärung: Männer verwenden eher nicht so häufig die Prestige-Variante [t] wie es Frauen tun. Dies sieht man z.B. unter anderem in Fig. 1 (mittlere und rechte Balken). Frauen bringen diese geschlechtsspezifischen Unterschiede den Kindern in CDS bei: d.h. sie wollen, dass Mädchen mehr wie Frauen, Jungen eher wie Männer klingen. Daher verwenden sie mehr [t] in CDS wenn sie mit Mädchen als wenn sie mit Jungen reden.

4. Es wird nicht ausführlich erklärt, weshalb in CDS [t] mit zunehmenden Alter des Kindes abnimmt. (Warum also nicht weiterhin den Kindern die Prestige-Variante beibringen?).

5. Es wird auch nicht erklärt, weshalb für finale /t/s, soviel [d] (Fig. 4, 5) in CDS und eigentlich mehr in CDS als in der Sprache mit Erwachsenen verwendet wird.

foulkes06.jop.pdf

+++++

Wie gesagt, ich würde nicht dieses Paper getrennt behandeln, sondern an geeigneten Stellen als Zusatzmaterial zu foulkes05.language.pdf hinzufügen.

Eventuell daher das wichtigste zu diesem Zweck:  
Für Erwachsene:

1. [d] wird häufiger von Männern als von Frauen verwendet
2. Fig. 2. Für das gleiche Alter und für das gleiche Geschlecht kommt [d] häufiger vor als [t]. z.B. [d] ist häufiger in wof als in mof; in wyf als in myf; in wym als in mym (Ausnahme: wom und wom sind ziemlich gleich).

Fig. 3 auf S. 414 sieht man

1. [d] wird häufiger in spontansprachlichen Äußerungen als in der Laborsprache verwendet.
2. Dies gilt vor allem innerhalb desselben Geschlechts und derselben sozialen Schicht. z.B. ist der Unterschied zwischen dem schwarzen und dem grauen Balken größer in der Mittelstandsklasse vs. Arbeiterschicht d.h. in mym als in wym; in myf als in wyf; in mom als in wom; der Unterschied zwischen dem schwarzen und dem grauen Balken ist auch meistens größer für Frauen als für Männer z.B.  
wyf > wym;  
myf > mym;  
wof > wom;  
nicht aber mof > mom)

Für Kinder (ab. S. 419)

.....

Fig. 9. Noch ein Bild um zu zeigen, dass Mütter in CDS häufiger [t] mit Mädchen als mit Jungen verwenden.

Fig. 10. ist nützlich, weil es zeigt, dass Kinder die feinen phonetischen Unterschiede der Muttersprache imitieren - d.h. je häufiger Mütter ihren /t/ präaspirieren, umso mehr wird diese Präaspiration von Kindern imitiert.

Fig. 11 eventuell auch, weil es zeigt wieder, dass die phonetischen Unterschiede zwischen Gender schon ab dem dritten Lebensjahr zum Vorschein kommen.

foulkes99.leedswip.pdf

+++++

der einzige Grund, weshalb dies nützlich ist, ist weil man die Sound-Dateien abspielen kann. Dadurch bekommt man gute Beispiele von Newcastle-Englisch. Der Rest des Papers bringt nichts neues, und kann weggelassen werden.

roberts97.childlang.pdf

+++++

(NB: MLU = mean length of utterance also Äußerungslänge).

Es geht hier wie in Foulkes et al um das Erlernen phonetischer Variation. Hier ist aber viel mehr die Frage, ob eher kategoriale phonologische Regel gelernt werden, oder ob viel einfacher Kinder individuelle Wörter imitieren.

16 Kinder wurden aufgenommen. Zusätzlich 8 der Mütter (S. 355). Es wurde untersucht, inwiefern /t, d/ in Konsonantenverbindungen und in finaler Position getilgt worden sind. Vor allem (S. 356)

[1] ob das morphemische Status des Wortes die Tilgung beeinflusst. In Erwachsenen gibt es einige Beweise, dass die Wahrscheinlichkeit der /t, d/ Tilgung eher vorkommt a > b > c

a. Monomorpheme. in monomorphemischen Wörtern wie 'nest'

b. Semi-weak. in polymorphemischen Wörtern, wo der Stamm sich in der Vergangenheit ändert (sleep->slept)

c. Weak past. in polymorphemischen Wörtern in Vergangenheitsformen, wo sich der Stamm nicht ändert (miss -> missed)

[2] ob der phonetische Kontext die Tilgung beeinflusst. Hier wird vermutet:

d > e > f > g > h

d: /t, d/ vor einem Obstruenten

e: /t, d/ vor einem Liquiden

f: /t, d/ vor /j, w/

g: /t, d/ vor einem Vokal

h: /t, d/ final vor einer Pause

zusätzlich: 'social class, gender, ethnicity, and conversational style have all been found to have an effect on (-t, d) deletion.'

Ergebnisse:

Fig. 1. Kinder tilgen /t, d/ auf eine ähnliche Weise wie es Erwachsene tun.

Figs. 2, 3 und der damit verbundene Text bringen nicht sehr viel mehr; die Abbildungen sind zu kompliziert. Ich würde sie weglassen.

Fig. 4. Im Allgemeinen:

Kinder und Erwachsene sind sich beide ähnlich, dadurch dass für beide Gruppen a > c

Es gibt aber zwischen Erwachsenen und Kindern in Fig. 4 deutliche Unterschiede. Hier sehen wir - im Gegensatz zu Erwachsenen - keinen Unterschied zwischen a, b.

Der Grund: Kinder behandeln Wörter wie 'slept' als wären sie monomorphemisch.

Es gibt zwischen b, c für Erwachsene keinen Unterschied - eventuell behandeln Erwachsene b, c ähnlich, weil sie beide polymorphemisch sind (d.h. Erwachsene aber nicht Kinder haben gelernt, dass b, c auf eine ähnliche Weise grammatikalisch strukturiert sind).

Dieses Ergebnis zeigt vermutlich, dass die Kinder nicht einfach Erwachsene imitieren - sondern dass sie \*auf eine andere Weise eine Regel anwenden\*. (S. 365)

----

Roberts (1997) zweifelt, ob die /t, d/-Tilgung soziophonetisch erklärt werden kann (S. 365). Der Grund ist ca. wie folgt. In einer früheren Untersuchung wurde gezeigt, dass Kinder oft /ɪn/ statt /ɪŋ/ ('ing') produzieren. Die Wahrscheinlichkeit der /ɪn/-Verwendung hing davon ab, mit welcher Gruppe Kinder sich unterhielten (mehr /ɪŋ/ mit Erwachsenen). Dies ist ein Beispiel der soziophonetischen Variation - da die Verwendung von /ɪŋ/ durch den Gesprächspartner bedingt ist (ähnlich wie 'style-shifting' in den Papers zu Oprah Winfrey - siehe #23). Aber es gibt keine Beweise, dass /t, d/-Tilgung auf eine ähnliche Weise soziophonetisch bedingt ist. Daher 'the present results suggest that in the case of (-t, d) deletion, it is the grammatical and phonological constraints which are first acquired by the children.'

Alter. Hier würde man wie in den Untersuchungen von Foulkes erwarten, wenn /t, d/-Tilgung soziophonetisch bedingt ist, dass Tilgung mit Alter verbunden ist. Das war nicht der Fall. Ebenfalls gab es keine Beziehung zwischen /t, d/ Tilgung und MLU (mean-utterance-length): 'This division, like the one by age, did not yield significant results in terms of overall probability of deletion' (S. 366).

Gender. 'It was found that there were, in fact, significant gender differences in deletion in these children. ...it was the girls who...were deleting (-t, d) at a higher rate than the boys (S. 367).'

Die Autorin hat aber wirklich keine Erklärung dafür.

Angenommen, dass die untersuchte Variante nicht von Lautwandel betroffen wird (=stable linguistic situation), dann würde man eher eine geringere Tilgung in Frauen (Mädchen) erwarten, da Frauen in solchen Fällen in ihrer Sprache konservativer sind. Sie sind nur innovativer, wenn die Variation mit Lautwandel verbunden ist (wenn die synchrone /t, d/-Tilgung sich später in Lautwandel umsetzt). Dann müsste man sagen (um dieses Ergebnis zu erklären, dass Mädchen häufiger /t, d/ tilgen), dass /t, d/ Tilgung die erste Phase eines Lautwandels ist, und es gibt keine Beweise dafür (keine Beweise, dass in 1950 'slep' statt 'slept' produziert wurde).

Vielleicht müsste man näher Wortschatzgröße und allgemeine Sprachkompetenz untersuchen. Es könnte sein (das spekuliert auch die Autorin), dass die Mädchen diesbezüglich fortgeschrittener sind. Das würde bedeuten, dass sie eher fähig wären, abstrakte Regeln wie /t, d/-Tilgung zu lernen, und das könnte dann dieses Ergebnis erklären.

## #27. The lexicon and language acquisition in early childhood

beckman07.pdf

+++++

Das Paper befasst sich mit der Interpretation im Rahmen der Exemplartheorie von Sprechfehlern produziert von zwei Sorten von Kindern mit Sprechstörungen.

Die ersten paar Seiten geben einen Überblick der relevanten Aspekte der Exemplartheorie. Die wichtigsten Teile davon:

1. Wörter sind eine Zusammensetzung der im Gedächtnis gespeicherten 'Spuren' (traces) des Wortes. d.h. jedes Mal, dass ein Wort gesprochen wird, hinterlässt es eine kontinuierliche, nicht segmentierte Spur des akustischen Signals im Gedächtnis. Diese Spur enthält viele phonetische Details - viel mehr als vorher vermutet worden sind - sowie auch andere Eigenschaften wie vom Sprecher.
2. Im Gedächtnis entstehen Dichtigkeiten/Knoten, wo sich die Spuren verschiedener Wörter überlappen. Diese Knoten ergeben die Phoneme oder eine endliche Anzahl phonologischer Einheiten der Sprache. Phonologische Einheiten sind also eine Abstraktion der (akustisch-phonetischen) Details in 1. Es ist wichtig hier zu verstehen, dass phonologische Einheiten nicht gegeben sind, sondern aufgrund der gespeicherten Spuren entstehen (insofern ist die Beziehung zwischen phonologischen Einheiten und Sprachsignalen unterschiedlich pro Person, da ja keine zwei Personen die selben Wortspuren gespeichert haben wird). Phonologische Einheiten d.h. phonologische Abstraktion ist notwendig, weil die Perzeption der Sprache wirkungsvoller ist, wenn in der Perzeption der Sprache die Wortspuren in endliche Einheiten zerlegt werden können - denn wenn Wörter im Lexikon für solche (phonologische) Einheiten kodiert sind, dann kann man im Lexikon Wörter schneller finden, die zu dem wahrgenommenen, produzierten Signal passen. Phonologische Abstraktion gibt auch Flexibilität: man kann schnell und problemlos Logatome (also neue Wörter) durch die Kombinatorik der phonologischen Einheiten bilden, auch wenn solche Reihenfolgen noch nie wahrgenommen worden sind.
3. Die Fähigkeit zur phonologischen Abstraktion entsteht im Laufe des Spracherwerbs - diese Fähigkeit geschieht nicht auf einmal, sondern ist zum Teil von der Größe des mentalen Lexikons abhängig: denn je mehr verschiedene Wortspuren wahrgenommen und im Gedächtnis gespeichert werden, umso wahrscheinlicher entstehen diese Knoten/Dichtigkeiten, die die Voraussetzung für die Phonologisierung (= Fähigkeit zur phonologischen Abstraktion/Entstehung phonologischer Einheiten) sind.
4. Es wird in diesem Paper vermutet, dass es zwei verschiedene Sorten von Sprechfehlern bei Kindern gibt.
  - a. SLI-Kinder (= children with specific language impairment, die eine Verzögerung in dem Erwerb verschiedener linguistischen Kompetenzen aufweisen), die vor allem Probleme mit 2. haben, also mit der Bildung phonologischer Einheiten/Abstraktion. Diese Gruppe ist Gegenstand der Untersuchung im 2. des Papers.

Diese SLI-Kinder müssten vor allem mit der Phonologisierung Probleme haben d.h. es müsste SLI-Kindern schwer fallen, neue Wörter zu bilden, die sie noch nicht gehört haben. Dies wird durch die Produktion von Logatomen mit häufigen und seltenen Phonemreihenfolgen i, Abschnitt 2. getestet. Diese Kinder müssten vor allem mehr Fehler als normale Kinder in der Erzeugung seltener Phonemreihenfolgen (die kaum gehört worden sind) zeigen.

b. PD-Kinder (= Kinder mit phonologischen Störungen). PD ist aber eine Fehlbezeichnung, denn laut den Autoren haben PD-Kinder nicht ein Problem mit 2. sondern eher ein Problem mit 1. (vielleicht auf eine ähnliche Weise wie bei Kindern mit leichten Hörstörungen). d.h. die auditive, unsegmentierte Spur pro wahrgenommenes Wort hinterlässt bei PD-Kindern (so die Hypothese) etwas ungenau die Details im Gedächtnis (also feine phonetische Details, genaue Information zum Sprecher werden im Gedächtnis nicht abgespeichert). Diese Kinder müssten größere Schwierigkeiten in der auditiven Diskriminierung im Vgl. zu normalen Kindern zeigen (z.B. größere Schwierigkeiten zwischen /p, t/ zu differenzieren. Zugleich müssten sie aber keine Probleme mit der Phonologisierung haben: im Gegensatz zu SLI-Kindern sollte es nicht der Fall sein, dass bei PD-Kindern mehr Sprechfehler in der Produktion seltener Phoneme vorkommen.

#### Abschnitt 2. vom Paper

.....

Es handelt sich um 16 Kinder mit SLI im Alter 8-13 Jahren. Solche Kinder haben auch im allgemeinen einen kleineren Wortschatz (daher auch mehr Probleme in der Phonologisierung - siehe 1. oben). Sie und normale Kinder mussten zwei verschiedene Sorten von Logatomen wiederholen:

mesəfəm (häufig)

gufegəd (selten)

beide sich auf die selbe Weise wortähnlich d.h. keine dieser Folgen verletzt phonotaktische Beschränkungen. Jedoch sind /me, sə/ häufig und /gu, fe/ viel seltener in englisch. [ə] ist ein r-gefärbtes Schwa, das im amerikanisch-englischen häufig auftritt.

#### Ergebnisse, Fig. 1

.....

Die zwei Bilder in Reihe 1 sind für die normalen Kinder (typically developing children). Die vertikale Achse ist in allen Fällen Wiederholungsgenauigkeit, also inwiefern die Kinder imstande waren, die Logatome erfolgreich zu erzeugen. Die offenen Vierecken und obere Regressionslinie sind für die häufigen, die geschlossenen Vierecken und untere Regressionslinie für die selteneren Logatome. Die horizontale Achse links ist Alter; die horizontale Achse rechts ist Expressive Vocabulary (was ich EV benennen werde) also eine Messung der Fähigkeit Wörter und verschiedene grammatikalische Kombinationen davon in der Produktion der Sprache zu verwenden.

Je älter das Kind, umso größer EV - daher ist es nicht erstaunlich, dass die Bilder oben links und rechts ähnliche Ergebnisse zeigen.

Die Ergebnisse zeigen (1) Je älter das Kind (je größer EV), umso weniger Fehler wurden in der Produktion der Logatome gemacht (auch nicht erstaunlich). (2) es gibt mehr

Fehler in der Produktion von seltenen als häufig vorkommenden phonemischen Reihenfolgen (die schwarzen Vierecken liegen unterhalb der nicht gefüllten Vierecken) und (3) ältere Kinder/Kinder mit größerem EV reagieren ähnlich auf häufige und seltene Phonemreihenfolgen (da die Regressionslinien rechts zusammenfallen d.h. häufig/selten hat keinen Einfluss auf die Anzahl der Fehler). Dies ist wichtig, weil es zeigt, dass ältere normale Kinder gelernt haben, Phoneme aus den phonetischen Details zu abstrahieren: sie sind also eher in der Lage, Phonemreihenfolgen, die sie eventuell nie/kaum gehört haben, trotzdem mit einer hohen Genauigkeit zu produzieren. Das ist möglich, weil ältere Kinder/Kinder mit hohem EV eher in der Lage sind, phonologische Einheiten miteinander zu kombinieren (jüngere Kinder können das nicht so gut, weil sie noch nicht so gut gelernt haben, phonologische Einheiten aus den phonetischen Details zu abstrahieren).

Es gibt noch ein Ergebnis (4): EV sagt die phonologische Genauigkeit doch etwas besser als Alter hervor: dies sieht man auch dadurch, dass die Punkte etwas näher an den Regressionlinien oben rechts vs. oben links liegen.

Jetzt zu den 2 Bildern unter rechts von Fig. 1

Es ist eine schlechte Abbildung, weil nicht alle Details klar ersichtlich sind.

Die Kreise sind hier die SLI also Kinder mit Sprachstörungen.

Die Vierecken sind normale Kinder \*im selben Alter\* (sodass Alter kein störender Faktor in der Interpretation der Ergebnisse ist).

Offene Symbole sind die häufigen, geschlossene Symbole die seltenen Phonemreihenfolgen.

Die horizontale Achse links ist EV, die horizontale Achse rechts ist eine numerische Verschlüsselung der Stärke der Sprachstörung.

Die Ergebnisse zeigen:

(a) SLI-Kinder haben einen kleineren EV (die Kreise in der Abbildung unten links sind links verschoben auf der horizontalen Achse)

(b) Je größer EV, umso weniger Fehler (das wussten wir schon aus (1) oben)

(c) Es gibt mehr Fehler in seltenen Phonemreihenfolgen (das wussten wir auch aus (3) oben)

(d) dies ist wichtig. Der Unterschied zwischen SLI- und normalen Kindern ist für seltene Phonemreihenfolgen ausgeprägter. Dies sieht man in der Abbildung unten links, weil die untere Regressionslinie steiler als die obere ist. Dies zeigt, dass SLI-Kinder eine größere Schwierigkeit als normale Kinder haben, phonologisch zu abstrahieren (also phonologische Reihenfolgen zu produzieren, die sie nur selten gehört haben).

Dieses Ergebnis erscheint noch deutlicher in der Abbildung unten rechts. Hier sieht man wieder, dass der Unterschied in der Genauigkeit Phonemreihenfolgen zu produzieren, für SLI-Kinder besonders ausgeprägt in seltenen Phonemreihenfolgen im Vgl. zu normalen Kindern ist (weil die Regressionslinie unten steiler ist als oben).

Abschnitte 3, 4

.....

Es handelt sich hier um einen Vergleich zwischen PD (Kinder mit 'phonologischen' Sprachstörungen) und TD (typically developing children, also normale Kinder ohne phonologische Störungen).

Das Ziel hier ist, zu zeigen, dass PD-Kinder im auditiven Bereich jedoch (im Gegensatz zu den SLI-Kindern oben) nicht im Bereich der phonologischen Abstraktion schlechter abschneiden. d.h. bei den PD-Kindern 'a good number of children with PD have subtle motor and/or perceptual deficits that affect their ability to form a robust encoding of the more successful articulatory patterns experienced in earlier babbling and word production'.

Einige Daten dazu in Fig. 2. Hier geht es um die Differenzierung zwischen (unmodiziertem) 'cap' vs 'cat' und 'tap' vs 'tack'. Die vertikale Achse ist eine Messung dafür, inwiefern Kinder zwischen diesen Wörtern diskriminieren können: je größer  $d'$ , umso besser die Diskriminierung. Die Daten aus 'cap' vs 'cat' und 'tap' vs 'tack' sind gemittelt, sodass es einen Daten-Punkt pro Kind gibt. Horizontale Achse links ist Alter, rechts ist eine Messung der phonologischen Sprachfähigkeit (also wie oft Kinder Fehler machen wie in Tabelle 2 des Papers). Kreise = PD, Vierecken = normale Kinder. Das wichtigste Ergebnis ist in Fig. 2 rechts: je schlimmer die Sprachstörung bei PD-Kindern, umso schwieriger fanden sie es, 'cap'/'cat' oder 'tap'/'tack' voneinander zu unterscheiden. Dies zeigt also (laut Autoren) ein Problem in der auditiven Diskriminierung zwischen Phonemen bei den PD-Kindern.

Das Ergebnis in Fig. 2 ist aber zweideutig. Dieses Ergebnis könnte entweder zustande kommen, weil PD-Kinder Probleme haben in dem Mapping von akustischem Signal auf Phoneme (= ein Problem der auditiven Diskriminierung wie oben) oder es könnte wie bei den SLI-Kindern sein, dass sie mit phonologischer Abstraktion Probleme haben.

Section 4 soll zwischen diesen Möglichkeiten unterscheiden.

Wenn es sich bei PD um eine Störung in der phonologischen Abstraktion handelt, dann müssten PD wie bei den SLI-Kindern Probleme in der Produktion seltener Phonemreihenfolgen haben.

Für dieses Experiment wurden etwas einfachere Logatome verwendet wie /bedag, donug/  
/ag/ ist häufig, und /ug/ ist selten.

Die Darstellung der Ergebnisse ist nicht sehr klar.

Die einzig wichtige Sache aus Fig. 3 ist die Abbildung unten rechts (alle anderen panels in der selben Abbildung können weggelassen/überflogen werden). Diese Abbildung in Fig. 3 unten rechts zeigt wie genau die Phonemreihenfolgen produziert wurden (vertikale Achse). Die horizontale Achse ist eine Messung der phonologischen Störung.

Die Hypothese, die hier getestet wird, ist wie folgt.

Wenn es eine Beziehung zwischen der Fähigkeit zur phonologischen Abstraktion und Häufigkeit der Sprechstörungen gibt, dann müssten PD-Kinder mehr Fehler bei den seltenen Reihenfolgen machen.

d.h.: die Regressionslinien in der Abb. unten rechts von Fig. 3 müssten auseinander gehen (die Regressionslinien sind für die häufigen (oben) und die seltenen Reihenfolgen).

Das war aber nicht der Fall. Daher 'the children with PD are no more affected by the probability of the target diphone sequence than are their age peers with TD. The result is in marked contrast to the result for the children with SLI.'

und auch daher:

'children with PD differ from children with TD in terms of the robustness of the encoding of word-forms in the parametric phonetic spaces, not at the level of the phonological grammar.'

#### Abschnitt 5 Schlussfolgerung

.....

Die Ergebnisse bestätigen also, dass es zwei Sorten von Störungen konsistent mit Exemplartheorie gibt: in der Fähigkeit zur phonologischen Abstraktion (SLI-Kinder) und in der genauen Wahrnehmung der hinterlassenen Wortspuren (PD-Kinder).

Dann einige nicht uninteressante Spekulationen für die Gründe weshalb es in der menschlichen Sprache diese zwei verschiedene Formen der Kodierung gibt.

1. Wegen der auditiven Kodierung von Wort-Exemplaren sind Aussprachen und soziale Kategorien - die Klassenzugehörigkeit z.B. - phonetisch so detailliert.

2a. Die viel grobere Phonologisierung, die auf diese phonetischen Details aufbaut, ist aber zugleich notwendig, um Wörter in der gesprochenen Sprache zu verstehen, d.h. um den salienten, bedeutungsentscheidenden Teil von den anderen phonetisch detaillierten Teilen des Sprachsignals zu trennen.

2b. Es wird auch vermutet, dass es ohne Phonologisierung kaum möglich wäre, für Kinder die Sprachperzeption mit ihrer eigenen Sprachproduktion zu verbinden. d.h. da ihre eigene Sprachproduktion notwendigerweise immer ganz anders im Vergleich zu den wahrgenommenen Signals ihrer Eltern/Betreuer ist (schon alleine wegen der unterschiedlichen Vokaltraktgröße), muss der Input zu ihrer Sprachproduktion (damit sie reden können) eine abstraktere Form (daher Phonologisierung) der Signale sein, die sie von den Eltern gehört haben.