

Verknüpfung digitaler Lemmalisten historischer Sprachstufen des Deutschen - wie und wozu?

- I Verknüpfung von Lemmalisten
- II Automatische Konstruktion mittelhochdeutscher Wortfamilien ausgehend vom Ahd.

I

Verknüpfung von Lemmalisten – wozu?

- Schlagendes Beispiel für den unmittelbaren Verknüpfungsnutzen: „Mittelhochdeutsche Wörterbücher im Verbund“
- Die verknüpften Lemmata bilden das Bindeglied zwischen lexikalischer Datenbeständen verschiedener historischer/ dialektaler Varietäten
- Sie ermöglichen Abfragen über Varietätengrenzen hinweg
- Sie ermöglichen einen Transfer von Informationen/ Annotationen von einem der verknüpften Datenbestände auf den andern

Verknüpfung von Lemmalisten – wie?

Die Möglichkeiten und Probleme einer weitgehend automatischen Verknüpfung sollen an zwei Beispielen thematisiert werden:

1. Althochdeutsch – Mittelhochdeutsch
2. Mittelhochdeutsch – Mittelniederdeutsch

I Verknüpfung von Lemmalisten

1. Althochdeutsch – Mittelhochdeutsch

Verknüpfung von ahd. und mhd. Lemmata: Ausgangsdaten

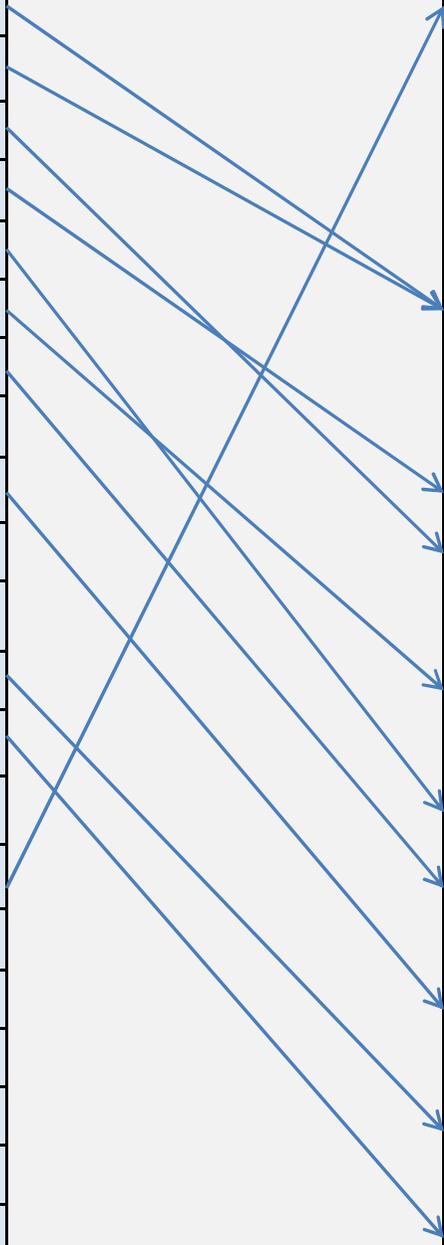
Ausgangspunkt sind digitalisierte Lemmalisten, und zwar

- der „Alphabetische Wortindex“ im II. Band von Jochen Spletts Ahd. Wb. (ca. 28.500 Lemmata)
- die MiGraKo-Lemmaliste (MiGraKo = Korpus der Mhd. Grammatik) mit ca. 73.000 Lemmata; sie beruht zwar zu einem Großteil auf Lemmalisten des neuen MWB, weicht aber in der Lemmagestalt aus grammatikspezifischen Gründen vielfach von den Wörterbuchansätzen ab: für die Lemmaverknüpfung ein Vorteil, wie sich noch zeigen wird.

Diese beiden Lemmalisten stehen zunächst unverbunden und im Einzelnen sehr verschieden nebeneinander:

<i>J.Splett, Alfab. Wortindex</i>
baldlîhho * Adv. → BALD
baldlîhhûn * Adv. → BALD
balda * st.F. → BALD
baldo (+*) Adv. → BALD
baldriân ** st.M. → Einzeleinträge B
baldên (*) sw.V. → BALD
balg (*) st.M. → BALG
balgheri * st.N. → BALG + HERI
balko (*) sw.M. → BALKO
balla (*) st.sw.F. → BAL
ballestar * st.N. → Einzeleinträge B
ballo * sw.M. → BAL
balo (+*) st.sw.M. → BALO
balolîhho * Adv. → BALO
balorât () st.M. → BALO + RÂTAN
balotât * st.F. → BALO + TUON

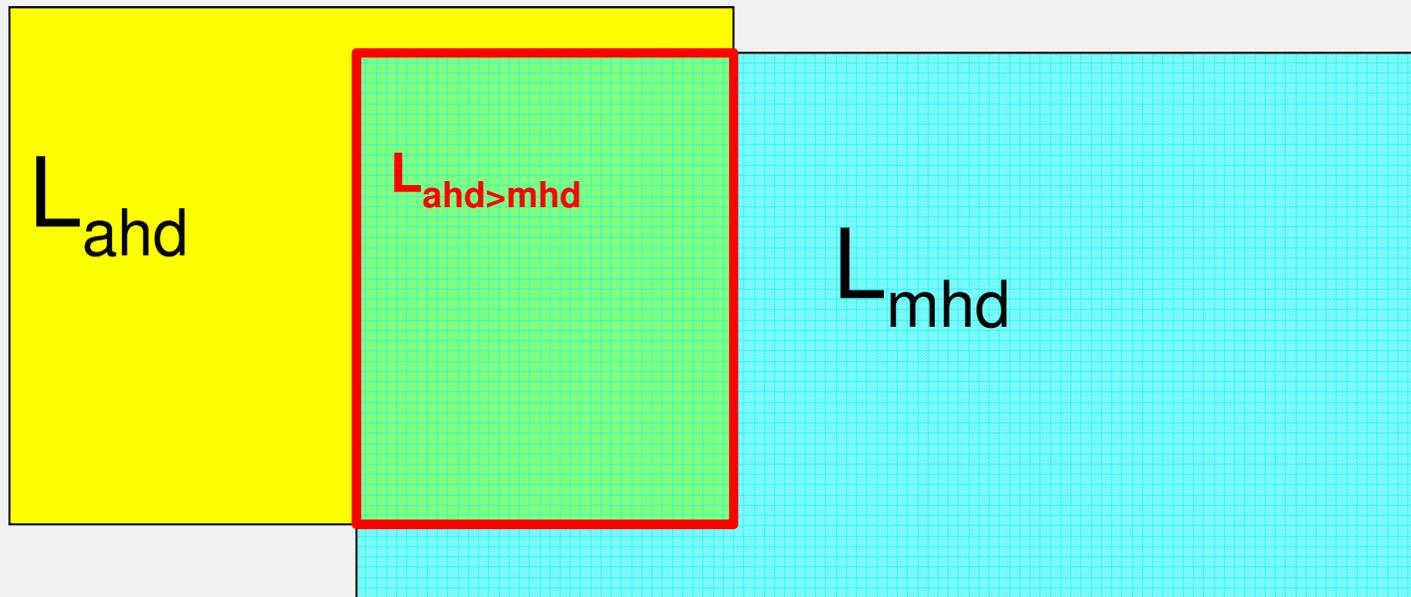
<i>MiGraKo-Lemmaliste</i>
bal-rât @m<u>
...
balas @m
bald @adj
...
bald-lîche @adv
bald-lôs @adj
...
balde @adv
balde @f
baldekîn @mn
balden @swv
...
baldriân @m
balg @m<u>
...
balke @m
...
ball(e) @m
...
balw @m



Vergleich der Lemmabestände von Ahd. und Mhd.

- Schon die beiden Listenausschnitte zeigen:
- Zu einem Teil der ahd. Lemmata gibt es mhd. Nachfolgeformen in der mhd. Liste, **baldlîhho** * Adv. – `bald-lîche @adv
- Es gibt aber auch ahd. Lemmata, die keine Entsprechung in der mhd. Liste haben, z.B. zu **balgheri** * st.N.
- Und es gibt noch weit mehr mhd. Lemmata, die keine Entsprechung in der ahd. Liste haben, z.B. `bald-lôs @adj. (Die meisten Fälle sind ausgelassen und durch ... angedeutet).

Gemeinsamer Lemmabestand von Ahd. und Mhd.



Nur der gemeinsame Bestand $L_{\text{ahd>mhd}}$ ist verknüpfbar, z. B.

baldlîhho * Adv. \Leftrightarrow `bald-lîche @adv

Rechtseindeutige Lemmaverknüpfung Ahd. → Mhd.

Zu einem verknüpfbaren ahd. Lemma gibt es in der Regel nur ein mhd. Lemma, mit dem die Verknüpfung möglich ist.

Für ein mhd. Lemma gibt es dagegen u. U. mehrere ahd. Verknüpfungspartner, z.B.:



Daher ist die Verknüpfung tunlichst von ahd. Seite her vorzunehmen.
Dafür spricht außerdem der weit kleinere ahd. Lemmabestand.

Verknüpfung von Lemmata: nach welchen Kriterien?

- Wann und warum verknüpfen wir Lemmata verschiedener Varietäten? Warum verknüpfen wir z.B. nhd. *warten* mit mhd. *warten* ‚schauen‘ und nicht mit *bîten* ‚warten‘?
- Weil für die Verknüpfung (\Leftrightarrow) üblicherweise nicht die Semantik entscheidend ist, sondern die genetische Verwandtschaft der Ausdrucksseite; kurz:
nhd. *warten* \Leftrightarrow mhd. *warten*, wenn nhd. *warten* < mhd. *warten*.
- Allgemein gilt also für ein ahd. Lemma *a* und ein mhd. Lemma *m*:
 $a \Leftrightarrow m$ genau dann, wenn $a > m$

Genetische Verknüpfung von Lemmata

- Die Bedeutung des Zeichens „>“ (bzw. „<“) ist nicht ganz klar: Meist = ‚durch regulären Sprachwandel geworden zu (bzw. entstanden aus)‘

So ist z.B. ahd. *firdriozan* durch regulären, „gesetzlichen“ Lautwandel (*io* > *ie*, *f-* > *v-* und Nebensilbenreduktion) zu mhd. *verdriezen* geworden.

Doch findet z.B. auch „ahd. *heffen* > mhd. *heben*“ oder „nhd. *schwören* < mhd. *swern*“, wo dies nicht zutrifft (analogische Umformung, nicht reguläres *ö* < *e*).

- Daher ist für das Folgende eine Präzisierung angebracht:

Genetische Verknüpfung von Lemmata

- $S_{\text{ahd} \rightarrow \text{mhd}}$ stehe für die Anwendung des gesamten **regulären** ausdrucksseitigen Sprachwandels vom Ahd. hin zum Mhd., z.B.

$$S_{\text{ahd} \rightarrow \text{mhd}}(\text{biscouwida}) = \text{beschöuwede}$$

Genauer: $S_{\text{ahd} \rightarrow \text{mhd}}$ ist eine Funktion, die ihrem Argument a einen Ausdruck zuweist, der aus a durch Anwendung des gesamten einschlägigen regulären Sprachwandels hervorgeht.

- Dann wird nicht immer, aber mit hoher Wahrscheinlichkeit gelten: wenn $a > m$, dann $m = S_{\text{ahd} \rightarrow \text{mhd}}(a)$

d.h.: wenn die ahd. Form eine mhd. Nachfolgeform hat, dann wird diese höchstwahrscheinlich das Ergebnis des regulären Sprachwandels sein.

So gilt z.B. für ahd. *firdriozan* :

$$firdriozan > verdriezen \text{ und } verdriezen = S_{\text{ahd} \rightarrow \text{mhd}}(firdriozan)$$

Genetische Verknüpfung von Lemmata

- Der enge Zusammenhang der Beziehung $>$ („wird zu“) und $S_{x \rightarrow y}$ lässt sich offenbar für unser Verknüpfungsproblem nutzen:
- Nehmen wir das ahd. Lemma *wînzurilâri* st.M. Gibt es dazu ein mhd. Nachfolgelemma m ?
- Wenn ja, dann ist nicht sicher vorhersagbar, wie m tatsächlich aussieht, wohl aber wie es **regulär aussehen müsste**, nämlich so:
$$m = S_{\text{ahd} \rightarrow \text{mhd}}(\text{wînzurilâri}) = \text{wînzürelære}$$
- Denn anders als die Beziehung $>$ ist $S_{\text{ahd} \rightarrow \text{mhd}}$ algorithmisierbar.
- Für ein Programm, das die ahd. Lemmaliste soweit wie möglich automatisch mit der mhd. Liste verknüpfen soll, ergeben sich also folgende Aufgaben:

Aufgaben eines Verknüpfungsprogramms

1. Zum ahd. Lemma a wird die mhd. Nachfolgeform $m = S_{\text{ahd}>\text{mhd}}(a)$ erzeugt.
2. Die mhd. Lemmaliste wird nach m durchsucht.
3. Je nach Suchergebnis wird folgendermaßen verfahren:
 - a. Wird m nicht gefunden, so bleibt a unverknüpft.
 - b. Wird m nur einmal gefunden, so wird die Verknüpfung mit a vollzogen.
 - c. Wird m mehrfach gefunden, so wird in allen Fällen eine Verknüpfung mit a vorgenommen, aber überall als fraglich markiert.

Anschließend müssen die als fraglich markierten Fälle manuell überprüft und entschieden werden.

Erzeugung des mhd. Lemmas: Überblick

1. Variantenerzeugung, z.B. *wuotag* → *wuotag* + *wuotîg* (⇔ mhd. *wüetig*)
2. Globale Ersetzungen, z.B. ahd. *ia, io* → *ie*; auch: *e* → *ë* (außer in *-en* sw.V.!)
3. Phonologisch(-morphologisch) konditionierte Ersetzungen (vor allem i-Umlaut)
4. Primär morphologisch gesteuerte Affixabtrennung + Nebensilbenreduktion
 - a. sukzessive Abtrennung und ggf. Umwandlung eines Suffixteils
 - b. ggf. Endsilbenreduktion des Restworts
 - c. sukzessive Abtrennung und ggf. Umwandlung eines Präfixteils
5. ggf. Bearbeitung von Komposita (Reduktion von Fugenvokalen)
6. Vereinigung von Präfixteil, Restwort und Suffixteil zum mhd. Lemma

Erzeugung des mhd. Lemmas: Überblick

1. Variantenerzeugung, z.B. *wuotag* → *wuotag* + *wuotîg* (⇔ mhd. *wüetig*)
2. Globale Ersetzungen, z.B. ahd. *ia, io* → *ie*); auch: *e* → *ë* (außer in *-en* sw.V.!)
3. Phonologisch(-morphologisch) konditionierte Ersetzungen (vor allem i-Umlaut)
4. Primär morphologisch gesteuerte Affixabtrennung + Nebensilbenreduktion
 - a. sukzessive Abtrennung und ggf. Umwandlung eines Suffixteils
 - b. ggf. Endsilbenreduktion des Restworts
 - c. sukzessive Abtrennung und ggf. Umwandlung eines Präfixteils
5. ggf. Bearbeitung von Komposita (Reduktion von Fugenvokalen)
6. Vereinigung von Präfixteil, Restwort und Suffixteil zum mhd. Lemma

Beispiele komplexer Präfix+Suffix-Bildungen:

biwollannussida

[4a] → biwollannuss[ede]

[4a] → biwollan[nissede]

[4b] → biwollen[nissede]

[4c] → [be]wollen[nissede]

gisezzida

[1] → gisetzida + gisezzida

[2] → gisëtzida + gisëzzida

[3] → gisètzida + gisèzzida

[4a] → gisètz[ede] + gisèzz[ede]

[4c] → [ge]sètz[ede] + [ge]sèzz[ede]

Erzeugung des mhd. Lemmas (1-2)

1. Variantenerzeugung, z.B.

Xag Adj. → (1) *Xag* + (2) *Xîg*, wenn *Xîg* nicht schon im Splett-Index

Xîg Adj. → (1) *Xîg* + (2) *Xag*, wenn *Xag* nicht schon im Splett-Index

Im Folgenden (3ff.) werden alle Varianten nacheinander abgearbeitet

<i>ahd. Lemma</i>		<i>Varianten</i>			<i>Lemmaregister</i>
wuotag	[2]	→ wuotag [3-4]	→	wuotig	----/---> —
		→ wuotîg [3-4]	→	wüètig	→ wüètig
sëzzen	[2]	→ sëzzen [3-4]	→	sèzzen	----/---> —
		→ sètzen [3-4]	→	sètzen	→ sètzen

2. Globale Ersetzungen, z.B. ahd. *ia, io* → *iè*, ahd. *hh, -h* → *ch* usw.
 bedingt: *e* → *ë*, außer in *-en* sw. V. / st. V.

Erzeugung des mhd. Lemmas (3)

3. Phonologisch(-morphologisch) konditionierte Ersetzungen, vor allem Durchführung von i-Umlaut:
 - a) vor *i*, *î* der Folgesilbe, z.B. (*suonlîh* [1] →) *suonlîch* [3a] → *süènlîch*;
hërmilîn → *hèrmilîn*
 - b) vor *-en*, wenn Wortart „V.“, z.B. *suonen* sw.V. [3b] → *süènen*
sëtzen sw. V. → *sètzen*

Erzeugung des mhd. Lemmas (4a)

4. Primär morphologisch gesteuerte Affixabtrennung + Nebensilbenreduktion

- a. wortartabhängige Abtrennung und ggf. Umwandlung von Suffixen, wenn das Restwort silbifizierbar bleibt, das trifft z.B. zu bei

(armil [2] →) èrmil → èrm-el,

aber nicht bei

*spil --/→ *sp-el; kindspil --/→ *kindsp-el [unzulässige Koda]*

bei erfolgreicher Suffixabtrennung wird 4a wiederholt, z.B.

hermilîn → hermil-în → herm-el-în [oder hermilîn → herm-elîn]

Die Adjektivsuffixe *-lîch* und *-ag/-îg* werden zu *-lich* bzw. *-ig* reduziert

Erzeugung des mhd. Lemmas (4b)

4. Primär morphologisch gesteuerte Affixabtrennung + Nebensilbenreduktion

b. ggf. Reduktion einer finalen Nebensilbe:

- alle kurzen Monophthonge im absoluten Auslaut → *e*
- *î* im absoluten Auslaut → *e*, wenn Restwort silbifizierbar
- *a, i, o, u* im durch einfachen Konsonanten gedeckten Auslaut → *e*
- *-ên/-ôn sw. V.* → *-en*

z.B. *boto* → *bote*, *rëgan* → *rëgen*, *salbôn sw.V.* → *salben*,
(*guotî* [3] →) *güètî* → *güète*, aber nicht auch *unfrî* → *unfre* [da *unfr*
nicht korrekt silbifizierbar]

Erzeugung des mhd. Lemmas (4c)

4. Primär morphologisch gesteuerte Affixabtrennung + Nebensilbenreduktion

- c. Abtrennung und ggf. Reduktion von Präfixen, wenn das Restwort so silbifizierbar ist, das sein erste Silbe keine Schwa-Silbe ist, z.B.

(gilônôn sw.V. [4a] →) gilôn-en [4b] → ge-lôn-en;

aber nicht:

*(ginôn sw.V. [4a] →) gin-en [4b] --/→ *ge-**n**-en*

Bei erfolgreicher Präfixabtrennung wird 4c wiederholt, z.B.

unfirholan [4b] → unfirholen [4c] → un-firholen [4c] → unver-holen

Erzeugung des mhd. Lemmas (5)

5. Ist der nach der Affixabtrennung verbleibende Kern ein- oder zweisilbig, so werden Präfixteil, Kern und Suffixteil zum fertigen mhd. Lemma zusammengefügt, das nun in der mhd. Lemmaliste gesucht werden kann.

6. Kompositum?

Ist der verbleibende Kern mehr als zweisilbig, so muss mit einem Kompositum mit Fugenvokal oder mit einem komplexen Kompositum gerechnet werden, also 1. Kompositionsglied mit Suffix(en) und/oder 2. Kompositionsglied mit Präfix(en).

Für diese Fälle würde sich ein semiautomatisches Vorgehen empfehlen.

Ergebnis 1: Verknüpfte Lemmaliste Althochdeutsch-Mittelhochdeutsch

	<i>Althochdeutsch</i>		<i>Mittelhochdeutsch</i>	<i>Ahd. Wortfamilie</i>
=	°bald-lîhho ÷Adv.	(*)	`bald-lîche @adv	BALD
=	°bald-lîhhûn ÷Adv.	*	`bald-lîche @adv	BALD
=	°balda ÷st.F.[(j)ô]	*	`balde @f	BALD
=	°baldo ÷Adv.	(+*)	`balde @adv	BALD
=	°baldriân ÷st.M.[a/i]	**	`baldriân @m	Einzeleinträge B
=	°baldên ÷sw.V.	(*)	`balden @swv	BALD
=	°balg ÷st.M.[a/i]	(*)	`balg @m<u>	BALG
/	°balg-heri ÷st.N.[ja]	*	*`balg-here @n	BALG + HERI
=	°balko ÷sw.M.[(j)an]	(*)	`balke @m	BALKO
/	°balla ÷st.sw.F.[(j)ô]/(j)ôn]	(*)	*`balle @f	BAL
=	°ballestar ÷st.N.[a]	*	`balster @mf n	Einzeleinträge B
=	°ballo ÷sw.M.[(j)an]	*	`ball(e) @m	BAL
=	°balo ÷st.sw.M.[wa]	(+*)	`balw @m	BALO
/	°balo-lîhho ÷Adv.	*	*`bale-lîche @adv	BALO
=	°balo-rât ÷st.M.[a/i]	()	`bal-rât @m<u>	BALO + RÂTAN
/	°balo-tât ÷st.F.[i]	*	*`bale-tât @f	BALO + TUON

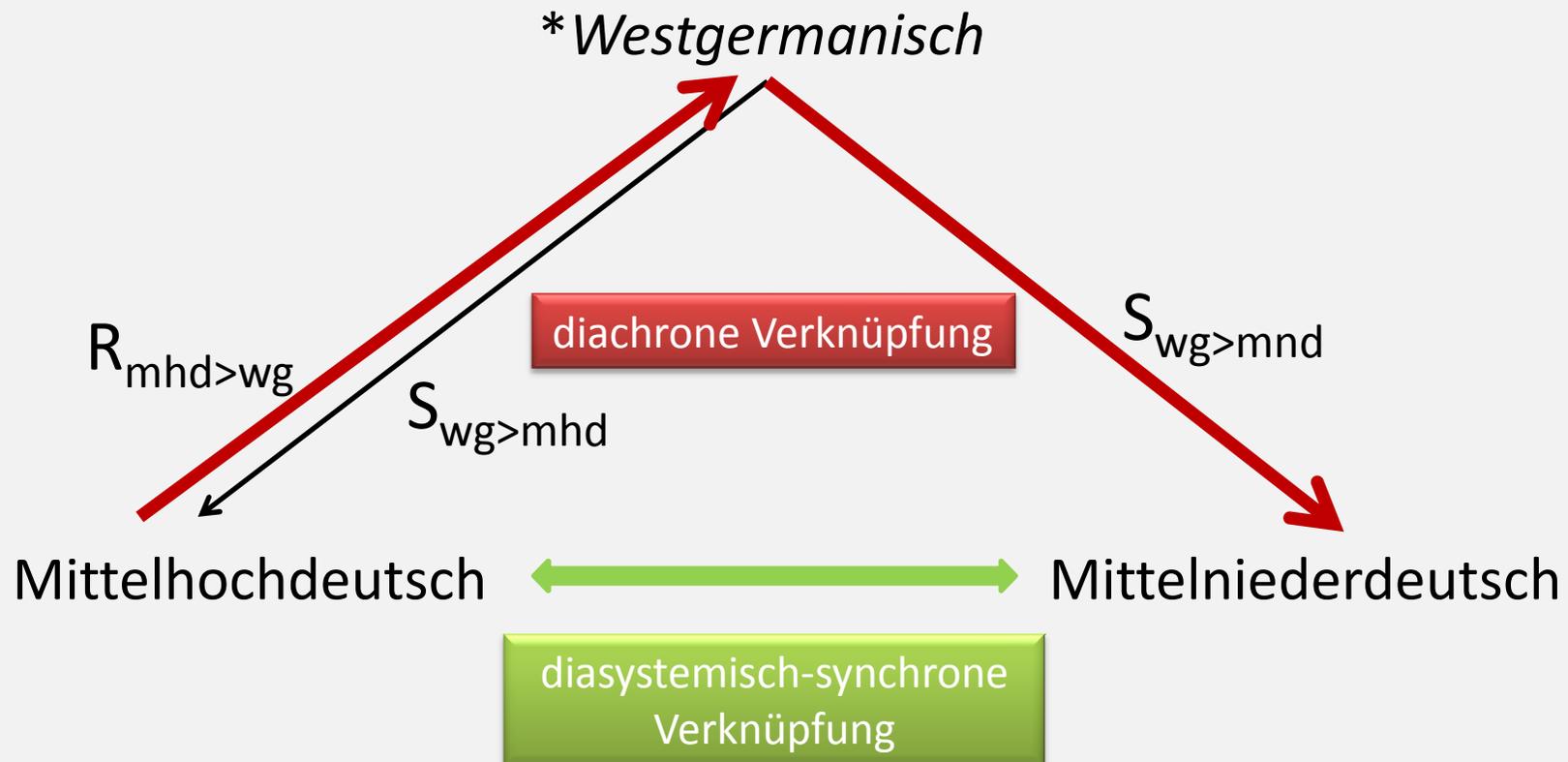
Ergebnis 2: Verknüpfte Lemmaliste Mittelhochdeutsch-Althochdeutsch

<i>MiGraKo</i>	<i>(Lexer/MWB)</i>	<i>MWB-ID</i>	<i>Althochdeutsch</i>	
`bal-rât @m<u> ^[]	``bal-rât #stm	:10836000	°balo-rât #st.M.	BALO + RâTAN
`bal-rèdære @m ^[]	``bal-reder #stm	:10842000	°-	
...
`balas @m ^[]	``balas #stm	:10683000	°-	
`bald @adj ^[]	``balt #a	:10959000	°bald #Adj.	BALD
`bald-hèit @f ^[]	``balt-heit #stf	:10965000	°-	
`bald-lich @adj ^[]	``balt-lich #a	:10968000	°bald-lîh #Adj.	BALD
`bald-lîche @adv ^[]	``balt-lîche #adv	:10974000	°bald-lîhho #Adv./ bald-lîhhûn #Adv.	BALD
`bald-lôs @adj ^[]	``balt-lôs #a	:10977000	°-	
`bald-spræche @adj ^[]	``balt-spræche #a	:10980000	°-	
`bald-spræche @f ^[]	``- #-	:-	°-	
`balde @adv ^[]	``balde #adv	:10704000	°baldo #Adv.	BALD
`balde @f ^[]	``balde #stf	:10707000	°balda #st.F.	BALD
`baldekîn @mn ^[]	``baldekîn #stm	:10719000	°-	
`balden @swv ^[]	``balden #swv	:10722000	°baldên #sw.V.	BALD

I Verknüpfung von Lemmalisten

2. Mittelhochdeutsch – Mittelniederdeutsch

Genetischer Zusammenhang Mhd. – Mnd. und Verknüpfungsweg



$S_{\text{wg}>\text{mhd}/\text{mnd}}$ = ausdrucksseitiger Sprachwandel vom Westgerm. zum Mhd./Mnd.

$R_{\text{mhd}<\text{wg}}$ = retrograder Sprachwandel/Rekonstruktion vom Mhd. zum Westgerm.

Lemmaverknüpfung Mhd. – Mnd. Überblick

- Die Verknüpfung sollte nicht auf diachronem Wege über das Westgerm. hergestellt werden, sondern direkt auf synchron-diasystemischem Weg vom Mhd. zum Mnd.
- Die Verknüpfung sollte vom Mhd. zum Mnd. hin erfolgen (Rechtseindeutigkeit der Verknüpfungsbeziehung).
- Besonders verknüpfungsg geeignet sind die Lemmata der MiGraKo-Liste und des Mnd. Handwörterbuchs.
- Der Verknüpfungsaufwand wird erheblich geringer sein als im Fall der ahd. – mhd. Lemmaliste.
- Ein besonderer Gewinn bestände darin, dass sich zugleich eine Verknüpfung mit der Lemma-ID des MWB ergeben würde.

Diachrone Lemmaverknüpfung Mhd. – Mnd.

Definition: Die Lemmata mhd. m und mnd. n sind miteinander verknüpfbar, wenn es ein westgerm. Lemma w gibt, sodass gilt:

$$n = S_{\text{wg}>\text{mnd}}(w) \text{ und } w = R_{\text{mhd}<\text{wg}}(m); \text{ daher: } n = S_{\text{wg}>\text{mnd}}(R_{\text{mhd}<\text{wg}}(m))$$

z.B. mhd. *bieten* – wgerm. **beodan* – mnd. *bêden*:

$$bêden = S_{\text{wg}>\text{mnd}}(*beodan) \text{ und } *beodan = R_{\text{mhd}<\text{wg}}(bieten); \text{ daher:}$$

$$bêden = S_{\text{wg}>\text{mnd}}(R_{\text{mhd}<\text{wg}}(bieten)) \text{ und folglich } bêden \Leftrightarrow bieten$$

Das gilt so allerdings nur für den „Erbwortschatz“, nicht aber für jüngere Lehnwörter, z.B. mhd. *prophête* – mnd. *profête*. Schon deshalb kann ein wgerm. „Hyperlemma“ keine Lösung sein.

Außerdem ist die rekonstruktive Erzeugung der wgerm. Form $w = R_{\text{mhd}<\text{wg}}(m)$ vielfach nicht unproblematisch.

Daher empfiehlt sich ein synchroner Verknüpfungsansatz:

Synchrone Lemmaverknüpfung Mhd. – Mnd.

Aufgrund der genetischen Verwandtschaft können auch synchrone Entsprechungsbeziehungen zwischen den Varietäten eines Diasystems ohne den Umweg über die gemeinsame wgerm. Vorstufe formuliert werden, z.B. in diasystemischen Formeln wie:

Mhd.	//	ie	uo	$\hat{o} \approx ou$...	$d \approx t$	//					
		—		≈	—	≈	—	≈	...	—	≈	...
Mnd. (Grammatik)	//	\hat{e}^4	\hat{o}^1	\hat{o}^2		d	//					
Mnd. Handwb.	//	\hat{e}	\hat{o}			d	//					

Während die Beziehung mhd. (MiGraKo-Liste) → mnd. in der Regel rechtseindeutig ist, gilt dies für die Beziehung mnd. → mhd. verschiedentlich nicht: Während z.B. ein mhd. *d* immer einem mnd. *d* entspricht, kann mnd. *d* sowohl mhd. *d* als auch *t* entsprechen, z.B. mnd. ¹*danne* ‚von dannen‘ ⇔ mhd. **d***anne(n)*, ²*danne* ‚Tanne‘ ⇔ mhd. **t***anne*; ähnlich mnd. (Handwb.) \hat{o} – mhd. *uo*, \hat{o} , *ou*, z.B. mnd. ¹*grôt* ‚Gruß‘ ⇔ **gruo***z*, ²*grôt* ‚groß‘ ⇔ **grô***z*

Synchrone Lemmaverknüpfung Mhd. – Mnd.

Bei der Beziehung mhd. (MiGraKo-Liste) → mnd. gibt es dagegen nur sehr beschränkte Störungen der Rechtseindeutigkeit:

1. Mhd. *t* entspricht zwar regelmäßig mnd. *d*, in wenigen Lexemen (mit germ. **tr*) jedoch mnd. *t*, z.B. in mnd. *trâch* ‚träge‘, *trân* ‚Träne‘, *trēden*, *trôst*, *trûwe*, *bitter*, *winter*; außerdem in diversen Lehnwörter (Lehnwörter sind generell der Hauptstörfaktor bei diasystemischen Entsprechungsregeln!).
2. Mhd. *ei* entspricht weit mehrheitlich mhd. \hat{e}^2 , selten aber auch \hat{e}^3 (vgl. Sarauw, Nd. Forschungen I, 145, 154ff; Lasch, Mnd. Gr. § 118-123)¹⁾ und dem nur vereinzelt vorkommenden mnd. *ei* (vgl. Lasch § 124ff; Sarauw I, 247ff).

Diese Regelausnahmen sind dem Verknüpfungsprogramm am besten durch eine einschlägige kleine Lemmaliste verfügbar zu machen.

1) Hier muss ich mich für eine Fehlinformation des Vortrags am 13.4. entschuldigen: Entgegen einer der gezeigten Folien unterscheidet das Mnd. Handwb. sehr wohl zwischen \hat{e}^2 (= \hat{e}^i , $\hat{e}(i)$) und \hat{e}^3 (= $\hat{e}i$)!

Synchrone Lemmaverknüpfung Mhd. – Mnd.

Inzwischen habe ich ein Programm geschrieben (Zeitaufwand: ca. 8 Stunden), das aus den mhd. MiGraKo-Lemmata mnd. Entsprechungslemmata erstellt; es arbeitet allerdings noch ohne Variantenerzeugung und berücksichtigt nur einzelne der bekanntermaßen abweichenden Kernwörter großer Wortfamilien wie z.B. mnd. *seggen* – mhd. *sagen*.

Die Ergebnisliste habe ich mit den ersten 100 Lemmata (ohne Verweise) des Buchstaben B im Mnd. Handwörterbuch verglichen. Es ist natürlich völlig offen, inwieweit diese kurze Alphabetstrecke überhaupt repräsentativ ist! Der Vergleich ergab Folgendes:

Nur 31 Lemmata sind mit mhd. Lemmata verknüpfbar; davon hat das Programm 19 (61,3%) korrekt erzeugt; dieser Anteil ließe sich mit Variantenerzeugung (z.B. *bak-* ~ *back-*) auf 25 (80,6%) steigern.

Ein Großteil der nicht verknüpfbaren Lemmata entfällt einerseits auf Komposita, z.B. mnd. *bāde-becken*, *-gast*, *-wāter*; andererseits auf Lehnwörter und ihre Ableitungen, so *baye* ‚Bucht u. Küste von Bourgneuf (Nantes) in Frankreich‘ und (fries.) *bâke(n)* ‚Zeichen‘

Mnd.-mhd. Lemmaliste (baalsche-bak)

	*mnd. (automatisch)	mhd. (MiGraKo-Liste)	Mnd. Handwb. (automatisch)	*mnd. (automatisch)	mhd. (MiGraKo-Liste)
baalsche <i>dênst</i>			bādeltît, f.		
babbelen, swv.			bādemāget, f.	bādemê't, f.	`bade-mèid @f
babbeler, m.			bādemôder, f.	bādemôder, f.	`bade-muoter @f
babbert			bāden, swv.	bāden, swv.	`baden @swv
bâch, m.	bâch, m.	`bâg @m	bādepenninc, m.	bādepenninc, m.	`bade-pfênning @m
bachante m.	bakant, m.	`bachant @m	bādequast, m.		
bachelken, porcellus			bāder, m.	bāder, m.	`badære @m
bachten, 1. adv, 2. prp.			bādermāget, f.		
bādebecken, n.			bāderôf, m/n.		
bādebōdene, f.	bādebüdde, f.	`bade-bütte @f	bādestê'n, m.	bādestê'n, m.	`bade-stèin @m
bādebrôk, f.			bādewāter, n.		
bādebûdel, m.			bādewerk, n.		
bādevat, n.	bādevat, n.	`bade-vazz @n	badewini		
bādegast, m.			bagelûn(e), f.		
bādegelt, n.	bādegelt, n.	`bade-gêlt @n	bâgen, swv.	bâgen, swv.	`bâgen @swv
bādehellinc, m.			bagge,		
bādehöike, m.			baggert, m.		
bādehôt, m.	bādehôt, m.	`bade-huot @n	bajâren,		
bādehûs, n.	bādehûs, n.	`bade-hûs @n	baye, f.		
bādekappe, f.	bādekappe, f.	`bade-kappe @f	bayevâr(e), m.		
bādekrût, n.			bāj(e)ken, n.		
bādekûven, n.			bayenvlôte, f.		
bādelāken, n.	bādelāken, n.	`bade-lachen @n	bayesch, adj.		
bādelôn, mn.	bādelôn, mn.	`bade-lôn @mn<(u)>	bayesolt, n.		
bādelstēde, f.	bādestat, f.	`bade-stat @f<u>	'bak, m. 'Rücken'		

= automatisch erzeugtes mnd. Lemma ist mit dem Lemma des Mnd.Handwb. identisch

= automatisch erzeugtes mnd. Lemma weicht vom Lemma des Mnd.Handwb. ab

Mnd.-mhd. Lemmaliste (²bak-bal)

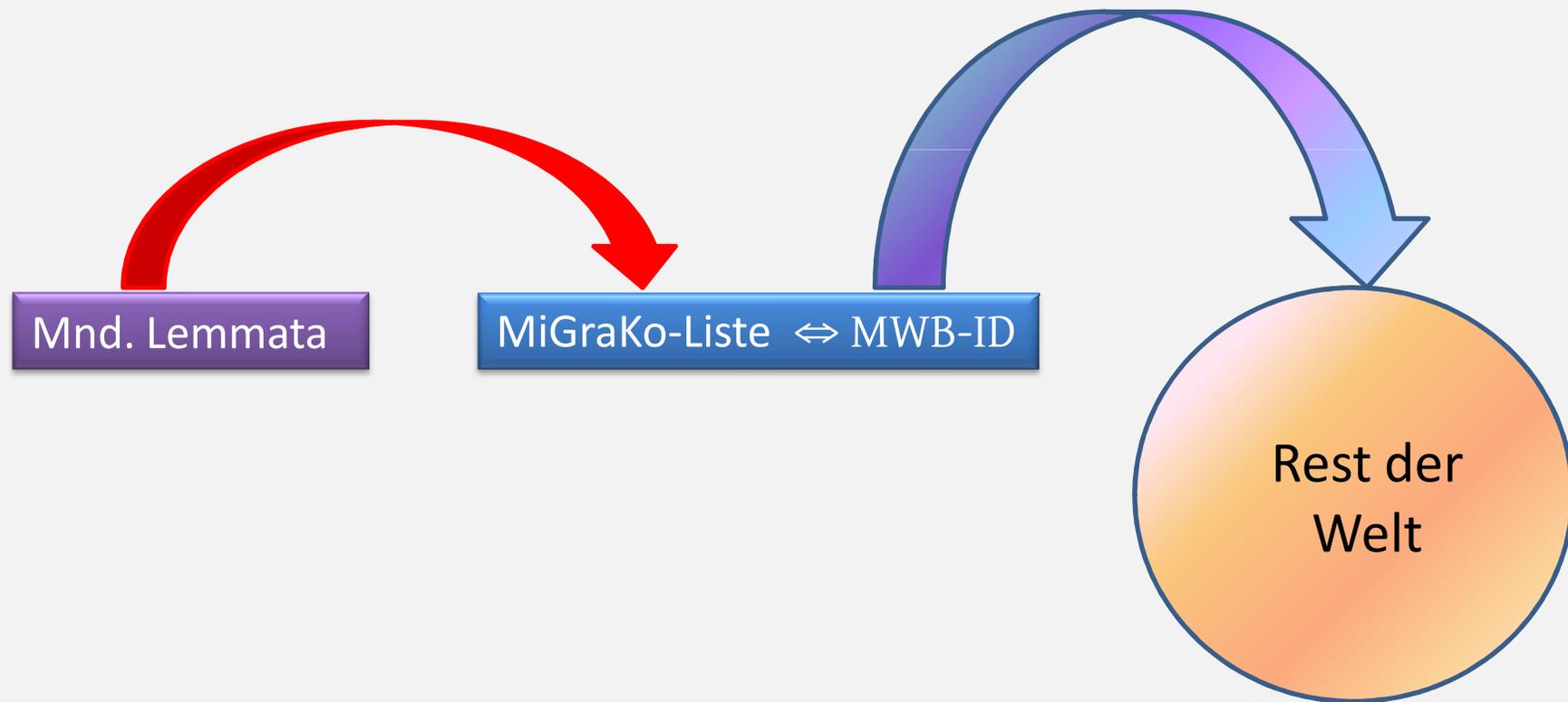
Mnd. Handwb.	*mnd. (automatisch)	mhd. (MiGraKo-Liste)	Mnd. Handwb.	*mnd. (automatisch)	mhd. (MiGraKo-Liste)
² bak, n. 'hölzernes Gefäß'			¹ backenkrût, n.		
³ (ge)bak, n. 'Gebackenes'	(ge)bak, m.	`back @m	² backenkrût, n.		
bakbört,			backenslach, m.	backenslach, m.	`backen-slag @m<u>
bakdarm, m.			backenstê'n, m.		
bāke, m.	bāke, m.	`bache @m	backe(ge)tûch, n.		
bâke, f.			backvisch, m.		
bākekêtel, m.			backhûs	bakhûs, n.	`bach-hûs @n
bâkemê'ister, m.			backîseren, n.		
bâken, swv.			backmê'ister, m.	bakmê'ister, m.	`bach-mèister @m
bâkengelt, n.			backôven, m.	bakôven, m.	`bach-oven @m
bâkenrecht, n.			backstê'n, m.		
bâkentolne, m.			backtê'iken, n.		
bākeren, swv.	bākeren, swv.	`bacher(e)n @swv	backtroch	bakdroch, m.	`bach-trog @m<(u)>
bakvanc, m.			backwerk, n.	backwerk, n.	`back-wêrk @n
bakharst,			baklôp, m.		
bakkalârius, m.			bakrêde, f.		
backammecht, n.			baksône, f.		
backdörnse, f.			bakstach, n.		
backe, f.	backe, m.	`back(e) @m	bakstêrt	bakstelte, f.	`bach-stêlze @f
(ge)backede, n.			bākswîn, n.		
backels(e), n.			bakwârd(e)s, adv		
backeltowe, n.			bakwascher, m.		
¹ backen, stv. ~ swv.	bāken, stv6.	`bachen @stv6	bakwôrdesch, adj.		
² backen, n.			bakwôrt, n.		
backenêl,			bal, m.	bal, m.	`ball @m

Lemmaverknüpfung Mhd. – Mnd.: wozu?

Eine Verknüpfung des mhd. und mnd. Lemmabestandes (die wennmöglich um eine Einbeziehung auch des frühnhd. Lemmabestandes zu erweitern wäre) erbrächte hauptsächlich folgenden Gewinn:

1. Einbezug des mnd. Wortschatzes, sofern er mit dem hochdeutschen (mhd.) genetisch verknüpfbar ist, in den lexikalisch recherchierbaren Teil des historischen deutschen Referenzkorpus (DDD).
2. Über die in der MiGraKo-Liste enthaltene Lemma-ID des Mittelhochdeutschen Wörterbuchs eine Anbindung an bereits bestehende und zukünftig auch weitere hochdeutsche Wörterbuchverbände.

Lemmaverknüpfung Mnd. – Mhd. und weiter



Lemmaverknüpfung Mhd. – Mnd.: wozu?

Eine Verknüpfung des mhd. und mnd. Lemmabestandes (die wennmöglich um eine Einbeziehung auch des frühnhd. Lemmabestandes zu erweitern wäre) erbrächte hauptsächlich folgenden Gewinn:

1. Einbezug des mnd. Wortschatzes, sofern er mit dem hochdeutschen (mhd.) genetisch verknüpfbar ist, in den lexikalisch recherchierbaren Teil des historischen deutschen Referenzkorpus (DDD).
2. Über die in der MiGraKo-Liste enthaltene Lemma-ID des Mittelhochdeutschen Wörterbuchs eine Anbindung an bereits bestehende und zukünftig auch weitere hochdeutsche Wörterbuchverbünde.
3. Der beschriebene Weg der halbautomatischen Verknüpfung über diasystemische Entsprechungsregeln verspricht einen wichtigen zusätzlichen Ertrag, nämlich
 - a) eine Explizierung und Präzisierung der diasystemischen Bezüge
 - b) die Möglichkeit, den Geltungsgrad der einzelnen Entsprechungsregeln statistisch zu ermitteln

II

Automatische Konstruktion
mittelhochdeutscher Wortfamilien
ausgehend vom Ahd.

Verknüpfte Lemmaliste Mittelhochdeutsch-Althochdeutsch

<i>MiGraKo</i>	<i>(Lexer/MWB)</i>	<i>MWB-ID</i>	<i>Althochdeutsch</i>	
`bal-rât @m<u> ^[]	``bal-rât #stm	:10836000	°balo-rât #st.M.	BALO + RâTAN
`bal-rèdære @m ^[]	``bal-reder #stm	:10842000	°-	
...
`balas @m ^[]	``balas #stm	:10683000	°-	
`bald @adj ^[]	``balt #a	:10959000	°bald #Adj.	BALD
`bald-hèit @f ^[]	``balt-heit #stf	:10965000	°-	
`bald-lich @adj ^[]	``balt-lich #a	:10968000	°bald-lîh #Adj.	BALD
`bald-lîche @adv ^[]	``balt-lîche #adv	:10974000	°bald-lîhho #Adv./ bald-lîhhûn #Adv.	BALD
`bald-lôs @adj ^[]	``balt-lôs #a	:10977000	°-	
`bald-spræche @adj ^[]	``balt-spræche #a	:10980000	°-	
`bald-spræche @f ^[]	``- #-	:-	°-	
`balde @adv ^[]	``balde #adv	:10704000	°baldo #Adv.	BALD
`balde @f ^[]	``balde #stf	:10707000	°balda #st.F.	BALD
`baldekîn @mn ^[]	``baldekîn #stm	:10719000	°-	
`balden @swv ^[]	``balden #swv	:10722000	°baldên #sw.V.	BALD

Konstruktionsverfahren – Überblick

1. Ermittlung aller verknüpften mhd. Lemmata mit demselben ahd. Wortfamilien-Kennwort (z.B. *BALD* oder Kernwort *KIOSAN*) = stabiler Kern der ahd. Wortfamilie
2. Analyse des ahd. Kernworts (Ablaut, Umlaut, etc.), z.B. *KIOSAN* → Ablautreihe IIb mit Gramm. Wechsel
3. Erzeugung aller möglichen mhd. Stammvarianten, z.B. *kies-*, *kôs-*, *kur-*, *kür-*, *kor-*, *kust-*, *kost-*
4. Ermittlung aller mhd. Lemmata, in denen (positionsbeschränkt) eine dieser Stammvarianten vorkommt = Menge aller potentiellen weiteren Familienmitglieder z.B. der Wortfamilie *KIESEN*
5. Manuelle Überprüfung insbesondere der Lemmata, die mehr als einer Familie zugeordnet worden sind.
6. Bearbeitung aller Fälle, die keiner bereits im Ahd. vorhandenen Wortfamilie zugewiesen werden konnten.
7. Ausgangspunkt dabei: die mhd. Simplizia, die im Ahd. noch nicht belegt sind; sie sind potentielle Kernwörter erst mhd. bezeugter Wortfamilien.
8. Weiter wie oben Punkt 3 – 5.

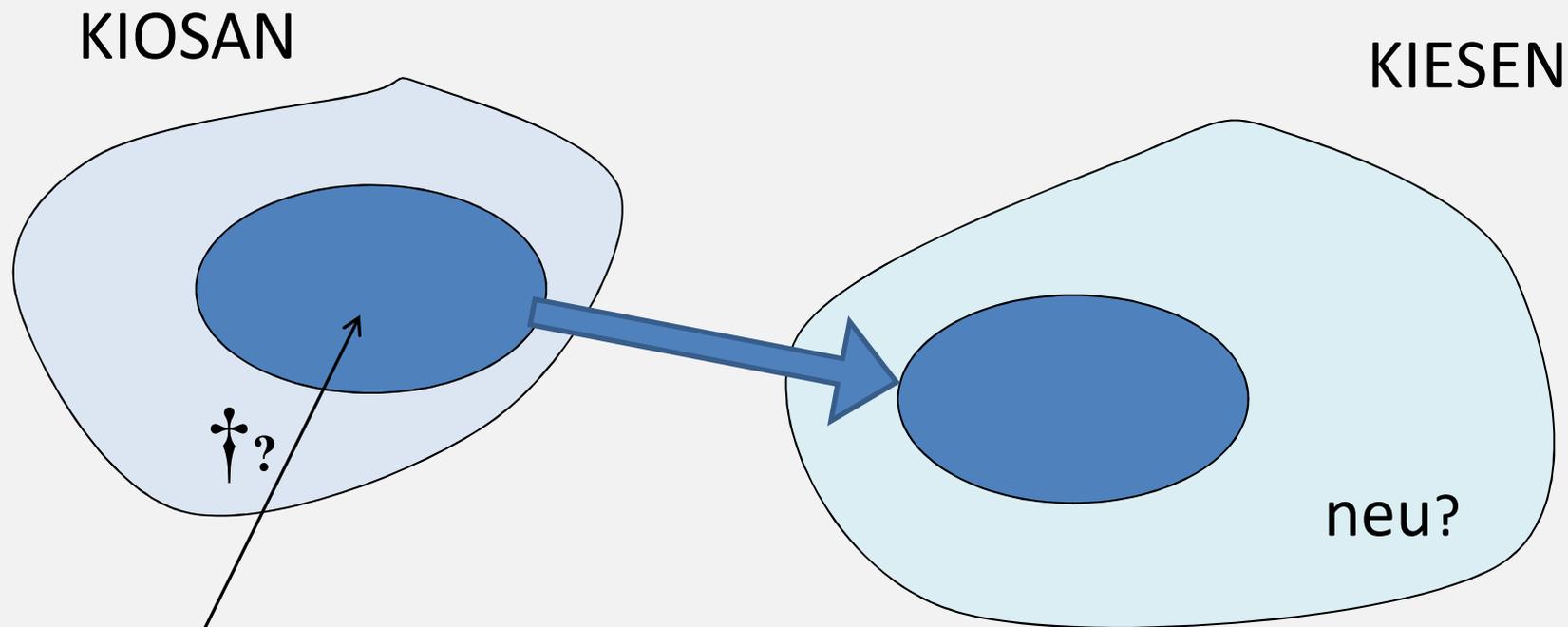
Konstruktionsverfahren (1)

1. Sortierung der mhd. Lemmaliste nach dem Feld „Kernwort der ahd. Wortfamilie“

Zusammensortiert werden so z.B. alle mhd. Lemmata mit dem ahd. *BALD* oder dem Kernwort *KIOSAN*.

Diese ahd./mhd. Lemmata bilden zusammen den bzgl. des Mhd. diachron stabilen Kern der ahd. Wortfamilie *KIOSAN*, d.h. die ahd. Mitglieder der Familie *KIOSAN*, die auch noch mhd. belegt sind:

Entwicklung der Wortfamilie ahd. KIOSAN /mhd. KIESEN



ahd. \Leftrightarrow mhd.: diachron stabiler Kern von ahd. KIOSAN

Konstruktionsverfahren (2)

2. Ermittlung des mhd. nicht mehr belegten Teils (†) der Familie ahd. *KIOSAN*: alle Lemmata des ahd. Lemmaindexes zum Kernwort *KIOSAN*, die in der MiGraKo-Liste keine Entsprechung haben.

Konstruktionsverfahren (3-4)

Ermittlung des im Ahd. noch nicht belegten Teils der Wortfamilie mhd. KIESEN

3. Analyse des Kernworts *KIOSAN*:

- *-an*, also starkes Verb
- *io + s* > mhd. *ie + s*, also Klasse IIb (Gramm.Wechsel *s* → *r* erwartbar)

4. Erzeugung aller möglichen mhd. Stammvarianten, im vorliegenden Fall ca. 13:

- *kies*, Umlaut → *kius*
- Ablaut → *kôs*, [mit analog. GW → *kôr*], i-Umlaut → *kæs*
- Ablaut [mit Beseitigung des GW] → *kus*, a-Umlaut → *kos*, i-Umlaut → *küs*
- ti-Abstraktum → *kust*, a-Umlaut → *kost*, i-Umlaut → *küst*
- Ablaut + GW → *kur*, a-Umlaut → *kor*, i-Umlaut → *kür*

Konstruktionsverfahren (5)

Ermittlung des im Ahd. noch nicht belegten Teils der Wortfamilie mhd. KIESEN

5. Durchsuchung des Lemmaregisters der mhd. Grammatik darauf, ob eine der Stammvarianten im jeweiligen Lemma in folgender Position vorkommt:

- linke Umgebung: Wortgrenze oder - (Segmentierungsstrich)
- rechte Umgebung: Wortgrenze, -, e, æ oder „altes Suffix“ (z.B. *unge*, *nisse*, *(e)de*, *ære* etc.)

Lemmata, die bereits in Schritt 1.-2. als Familienmitglieder ermittelt wurden, werden dabei übergangen; ebenso alle Lemmata, die im Lemmaregister bereits mit einer anderen ahd. Wortfamilie verknüpft sind, wie z.B. mhd. *kôse* ‚Rede‘, *kôsen* ‚sprechen, plaudern‘ mit der Wortfamilie *KÔSA* (< lat. *causa*) oder *kôr* ‚Chor‘ mit *KÔR*.

Konstruktionsverfahren (6)

Ermittlung des im Ahd. noch nicht belegten Teils der Wortfamilie mhd. KIESEN

6. Das Resultat dieses Suchvorgangs ist dann die Menge aller weiteren potentiellen Familienmitglieder, wenn die vorausgehenden Schritte „korrekt“ erfolgt sind.

Allerdings wird nicht jedes potentielle Mitglied auch tatsächlich zur Familie gehören, so z.B. zwar *kosten* ‚(aus)probieren, prüfen, etc.‘, aber nicht auch *kosten* ‚einen bestimmten Preis haben, ausgeben, beköstigen‘ (< lat. *cōnstāre* ‚zu stehen kommen‘), das zu einer separaten erst mhd. Wortfamilie gehört.

Daher ist eine abschließende manuelle Nachprüfung angezeigt.

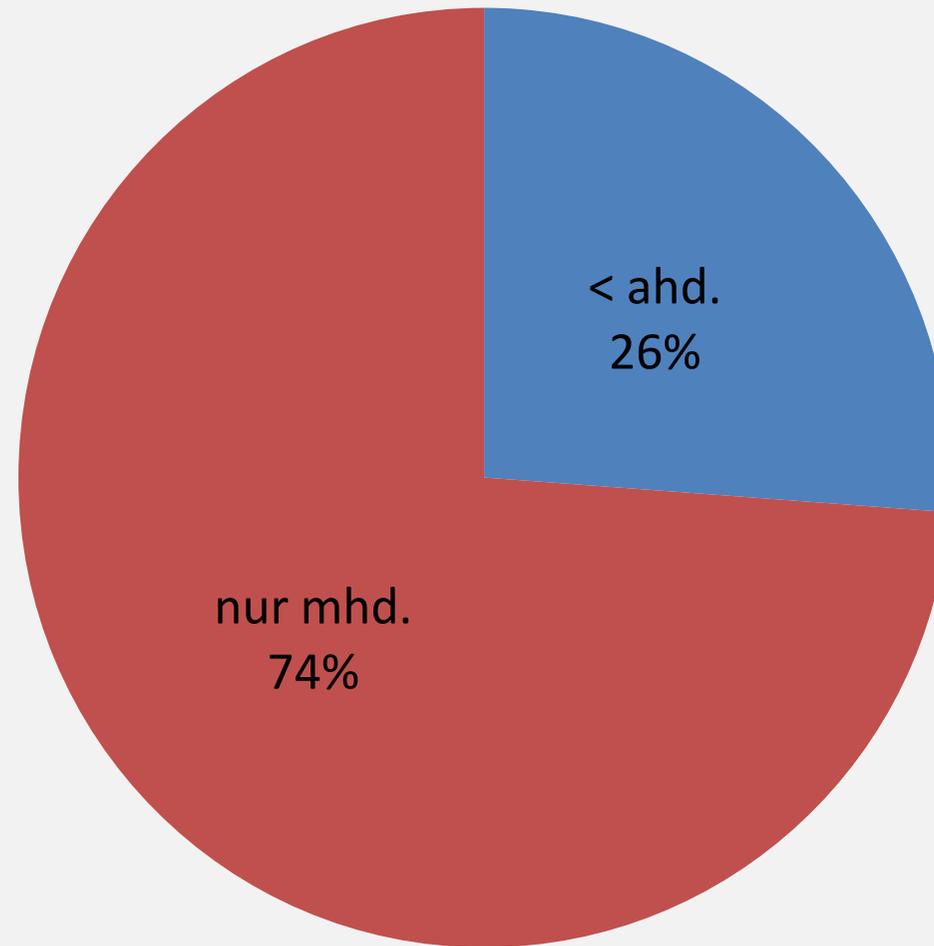
Wozu?

- Bezogen auf das Ahd. lassen sich etwa folgende Fragen untersuchen:
 - Wie verhält sich der stabile Kern zum Gesamt der einzelnen Wortfamilien?
 - Was besagt das über die unterschiedliche „Vitalität“ der Wortfamilien?
 - Gibt es jeweils Unterschiede hinsichtlich der Bildungstypen?

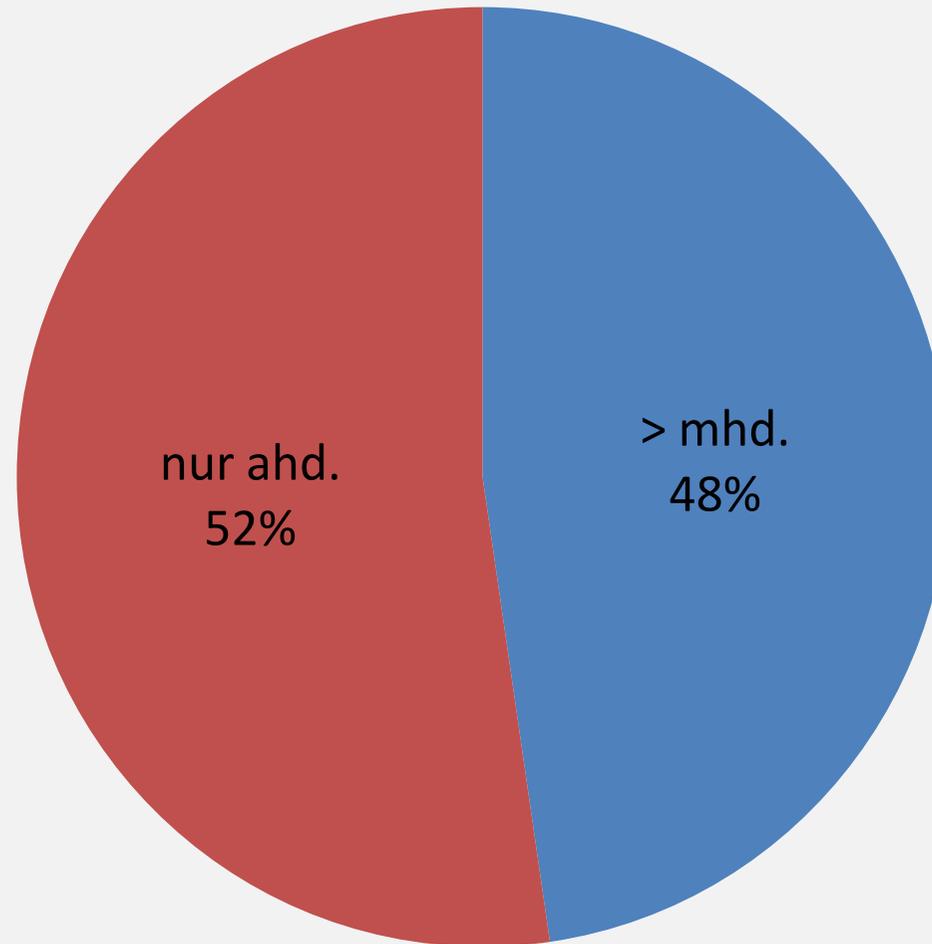
Wozu?

- Bezogen auf das Mhd. :
 - Welche Wortfamilien hatten in mhd. Zeit überdurchschnittlichen, welche unterdurchschnittlichen Zuwachs, welche sind u.U. geschrumpft?
 - Was besagt das für Produktivitätsaspekte der Wortbildung im Mhd.?
 - Welche Wortfamilien sind im Mhd. neu hinzugekommen?

Wortfamilie mhd. KIESEN

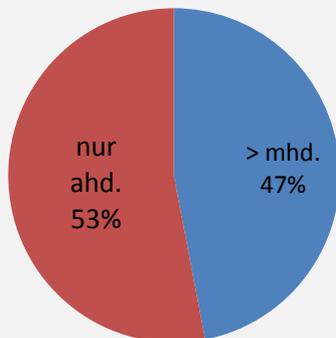


Wortfamilie ahd. KIOSAN

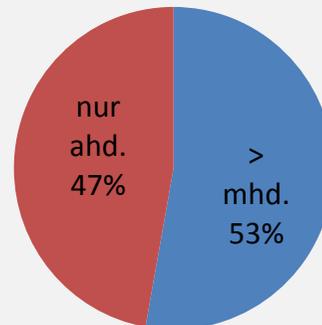


Ahd.

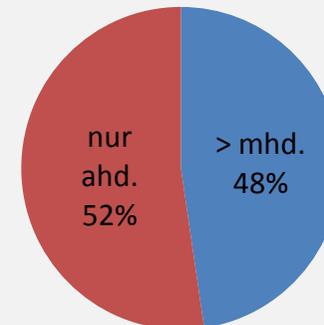
Gesamtliste



**Wortfamilie ahd.
BIOTAN**

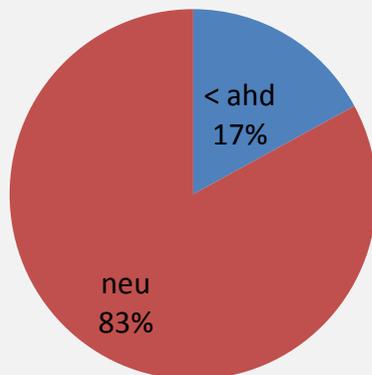


**Wortfamilie ahd.
KIOSAN**

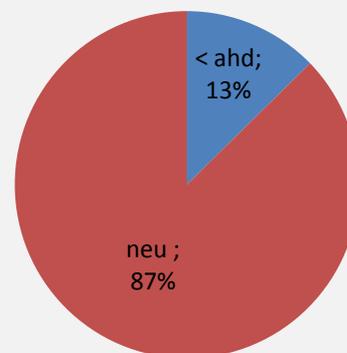


Mhd.

Gesamtliste



BIETEN



KIESEN

