

# Distant Reading den Zusammenhang zwischen Form und Inhalt in der mittelhochdeutschen Dichtung

Christopher Hench  
PhD Candidate, UC Berkeley

# Morphologie des Texts

- Moretti, *Distant Reading*
  - “Here is a modest example of what quantitative stylistics could do: take those units of language that are so frequent we hardly notice them and show how powerfully they contribute to the construction of meaning.”

# Zwei Methoden

1. Lyrische Klanglandschaften und Inhalt
2. Korrelation zweierlei Clusteranalysen (auf formale Merkmale sowie auf lexicalische Merkmale)

# Klanglandschaften

Dêst ein nôt, daz mich ein man  
vor al der werlte , swes er wil  
sol ich, des ich niht enkan,  
beginnen, daz ist mir ein swaerez spil  
Ich hât ie vil staeten muot  
nu muoz ich leben als ein wîp,  
diu minnet und daz angestlîchen tuot

– Reinmar *MF* (192,25–214C)

Swenne ich sî mit mîner valschen rede betrûge,  
sô het ich sî unreht erkant  
und gevâhe sî mich iemer an deheiner lûge,  
sâ sô schupfe mich zehant  
Und geloube niemer mîner klage,  
dar zuo niht, des ich sage  
dâ vor mûeze mich got behûeten alle tage

– Reinmar *MF* (173,13–96C,53b)

Dêst ein nôt, daz mich ein man  
vor al der werlte twinget, swes er wil  
sol ich, des ich niht enkan,  
beginnen, daz ist mir ein swaerez spil  
Ich hât ie vil staeten muot  
nu muoz ich leben als ein wîp,  
diu minnet und daz angestlîchen tuot

– Reinmar *MF* (192,25–214C)

9/59 = 15.25%

Swenne ich sî mit mîner valschen rede betrûge,  
sô het ich sî unreht erkant  
und gevâhe sî mich iemer an deheiner lûge,  
sâ sô schupfe mich zehant  
Und geloube niemer mîner klage,  
dar zuo niht, des ich sage  
dâ vor mûeze mich got behûeten alle tage

– Reinmar *MF* (173,13–96C,53b)

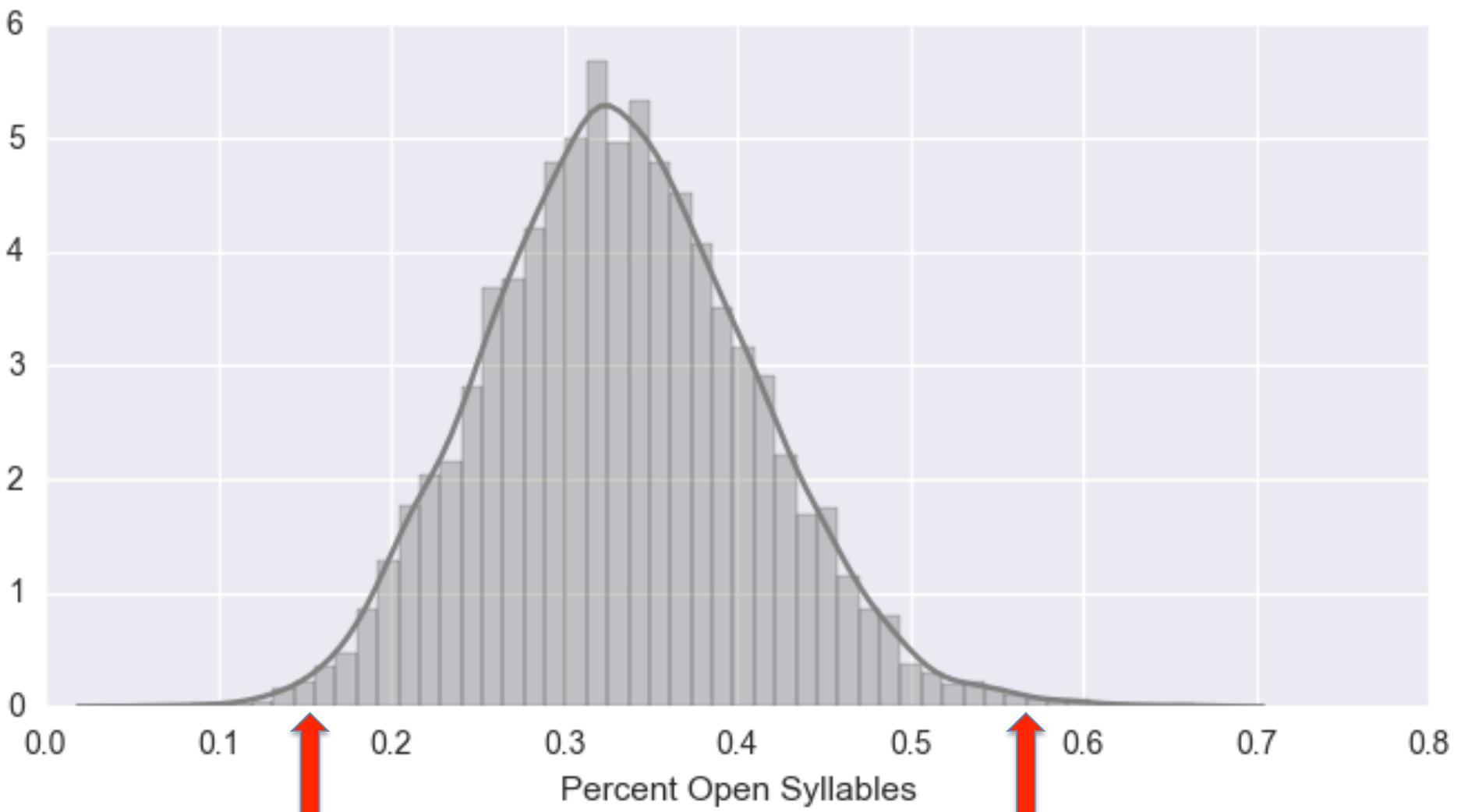
42/73 = 57.53%

---

	stanzas
count	7856
median syllables per stanza	81
median lines per stanza	9
median syllables per line	8
median percent open syllables	32.88%

---

Table 1: Summary statistics for soundscape stanza analysis



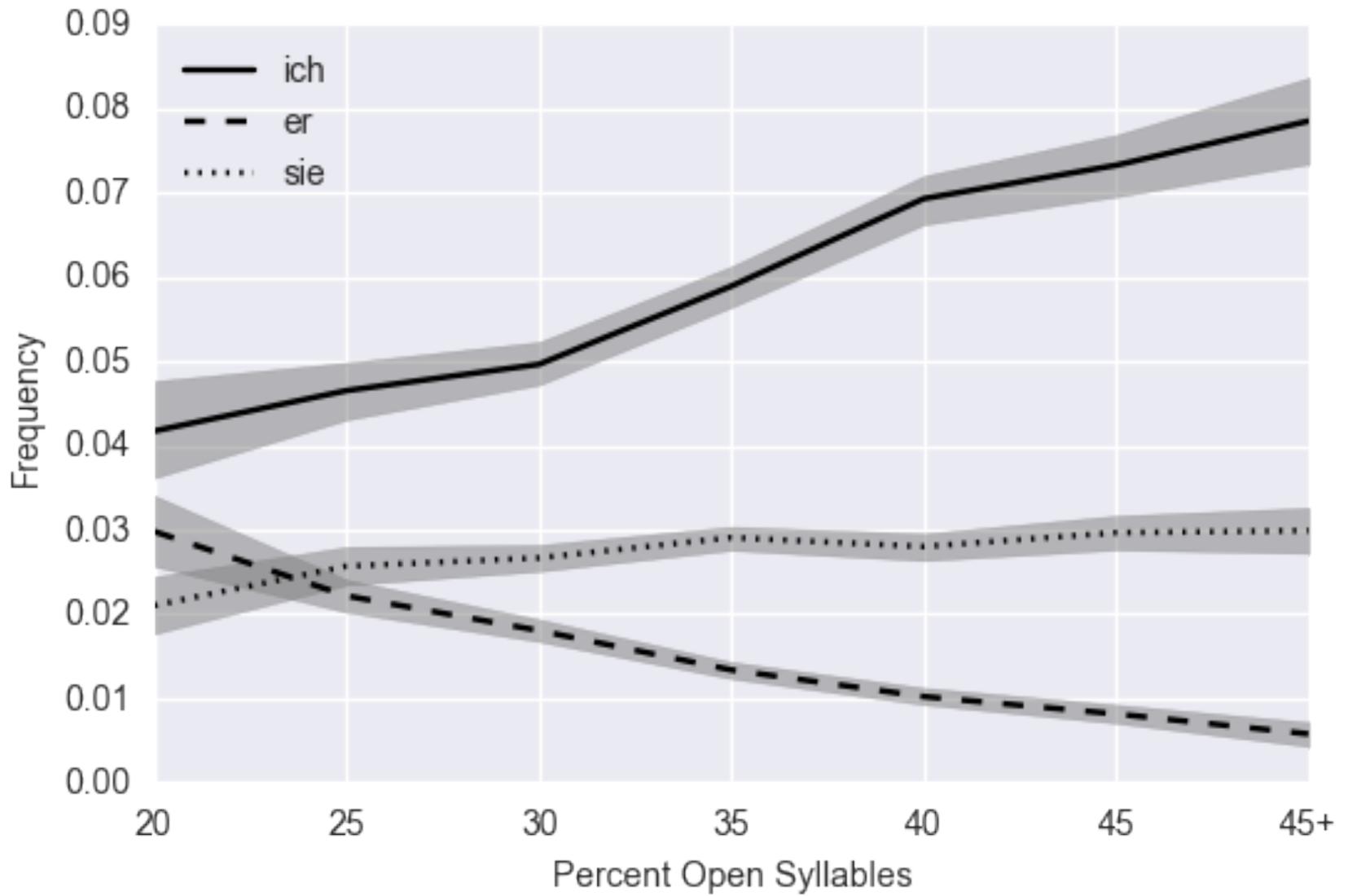
Dêst ein nôt

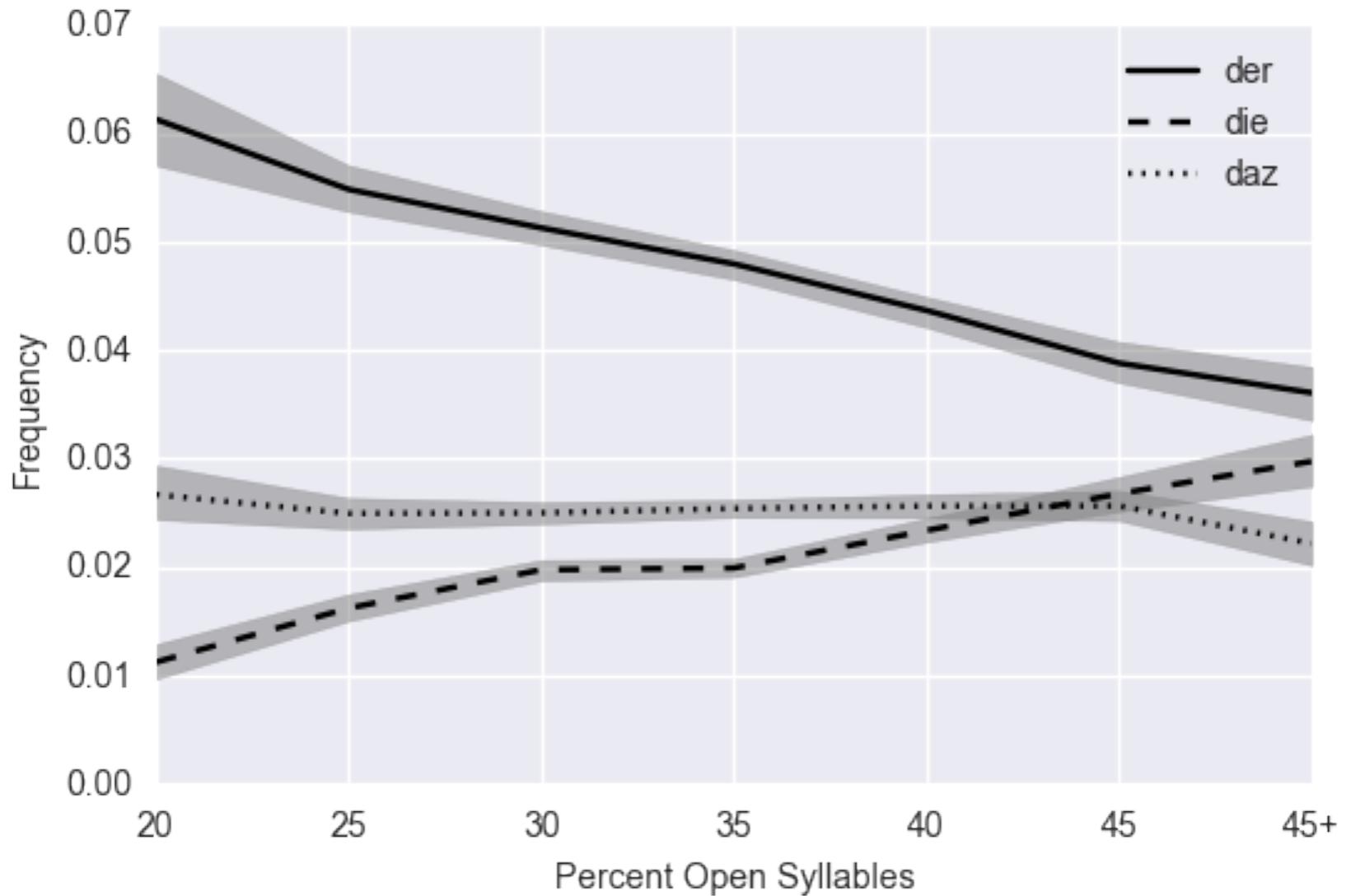
Swenne ich sî

# Methode

---

percent open	stanzas
< 20%	272
< 25%	803
< 30%	1582
< 35%	2055
< 40%	1592
< 45%	974
<u>≥ 45%</u>	516

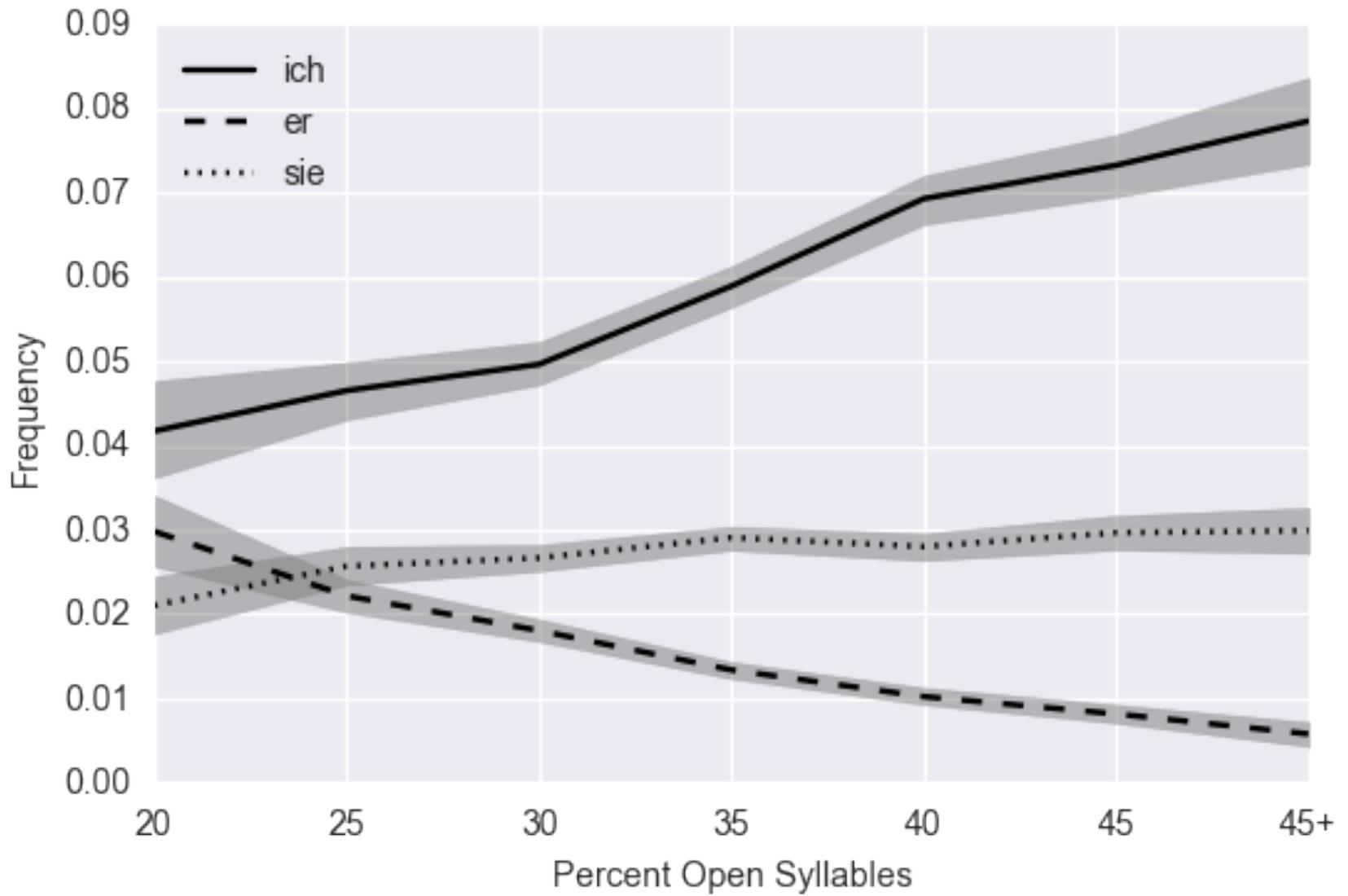




# Handschriften und Ausgaben

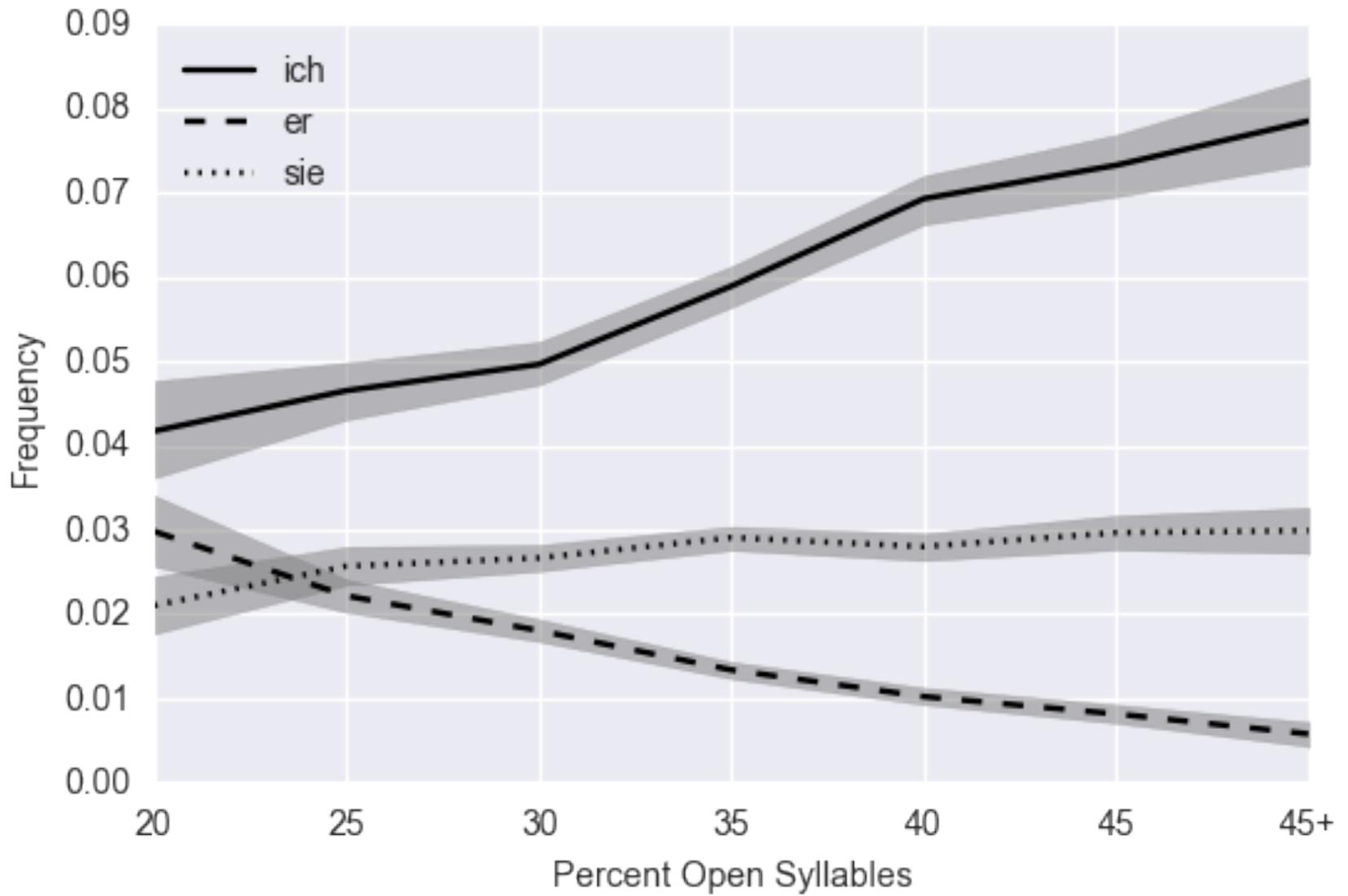
	stanzas	sound diff.
Dietmar von Aist	42	.79% $\pm$ 1.00 %
Rubin	68	.82% $\pm$ .89%
Der wilde Alexander	35	1.09% $\pm$ .99%
lyric total	145	.86% $\pm$ .95%

# Gattungen und Dichter



Owê,  
daz **er** sô dicke sich  
bî mir ersehen hât!  
als **er** endahte mich,  
sô wolt **er** sunder wât  
mîn arme schouwen blôz.  
ez was ein wunder grôz,  
daz **in** des nie verdrôz.  
dô tagte ez.

Heinrich von Morungen *MF* 143,22–93C ff



**Ich** solt aber dur süezen  
grüezen meien walt heid ouwe  
und der kleinen vogel süezez singen,  
lieze eht **mir** an ir gelingen  
trût **mîn** trôst, **mîs** herzen frouwe,  
daz si **mînen** kumber wolde :  
seht, sô wurde **ich** noch an fröiden rîche.  
truter lip, nu tuot genædecliche:  
rôter munt, du maht **mîn** leit verdringen. (35.71% offen, ich = 12.72% Lemmata)

Süeziu minne, sît dîn minne  
sinne krenket zallen stunden,  
wie sold **ich** dan iemer frô belîben?  
lieber lîp vor allen lîben,  
heilet **mir mîns** herzen wunden,  
daz **mîn** fröide lige an dem gewinne.  
tuot ir daz, sô wirde **ich** froidebære.  
sælic wîp, nu scheidet **mich** von swære.  
Minne, dû maht **mir mîn** leit  
vertrîben. (39.29% offen, ich = 14.81% Lemmata)

**Ich** hân fröide von ir eine:  
seine troestet **mich** ir güete.  
dâ von muoz **mir** spilndiu fröide swinden.  
lieze sie **mich** gnâde vinden,  
seht, so fröite **mîn** gemüete,  
daz **mir** wurde ir rôter kus, ein kleine;  
so wær **ich** vil manger sorgen âne.  
triutelehter lîp, **ich** lebe in wâne  
daz **ich** fröide von iu müge enpfinden.  
(49.41% offen, ich = 16.07% Lemmata)

# Korrelation der Clusteranalysen

# Einfache formale Morphologie des Texts

- Verslänge (Worte, Silben)
- Reim
- Silbenqualitäten (offen, geschlossen)

Versmaß, Reimstruktur, Melodie, Klang (Kragl, 2011)

# Codierung der Merkmale

Ist zwîvel herzen nâchgebûr

C - O C - C C - C O C - X

daz muoz der sêle werden sûr

C - C - C - OO - C C - C - 1

# 10-gram Ausschnitte

Ist zwîvel herzen nâchgebûr

C - O C - C C - C O C - X

- C-OC-CC-CO, -OC-CC-COC, OC-CC-COC, C-CCCOC-X

daz muoz der sêle werden sûr

C - C - C - OO - C C - C - 1

- C-C-C-OO-C, -C-C-OO-CC, C-C-OO-CC-, -C-OO-CC-C, C-OO-CC-C-, -OO-CC-C-X

*Parzival*:<sup>49</sup>

der tac gein dem âben—de zôch,

→ dô kom er an ein wazzer snel,

→ daz was von sînem duzze hel,

ez — gâben die velse ein ander.<sup>50</sup>

*Willehalm*:<sup>51</sup>

diu stuont uf, mit zuht — si sprach,

→ e daz sich schiet der vürsten rat:

→ <swer zuht mit triuwen hinne hat,

der — ruoche hoeren miniu wort.<sup>52</sup>

sequence	share in corpus
C-C-C-C-C-	0.006201
-C-C-C-C-C	0.004032
C-XC-C-C-C	0.003100
C-1C-C-C-C	0.002960
C-C-C-OC-C	0.002868
-C-C-C-OC-	0.002754
-C-XC-C-C-	0.002726
-C-C-OC-C-	0.002659
C-C-C-C-OC	0.002570
C-C-XC-C-C	0.002439

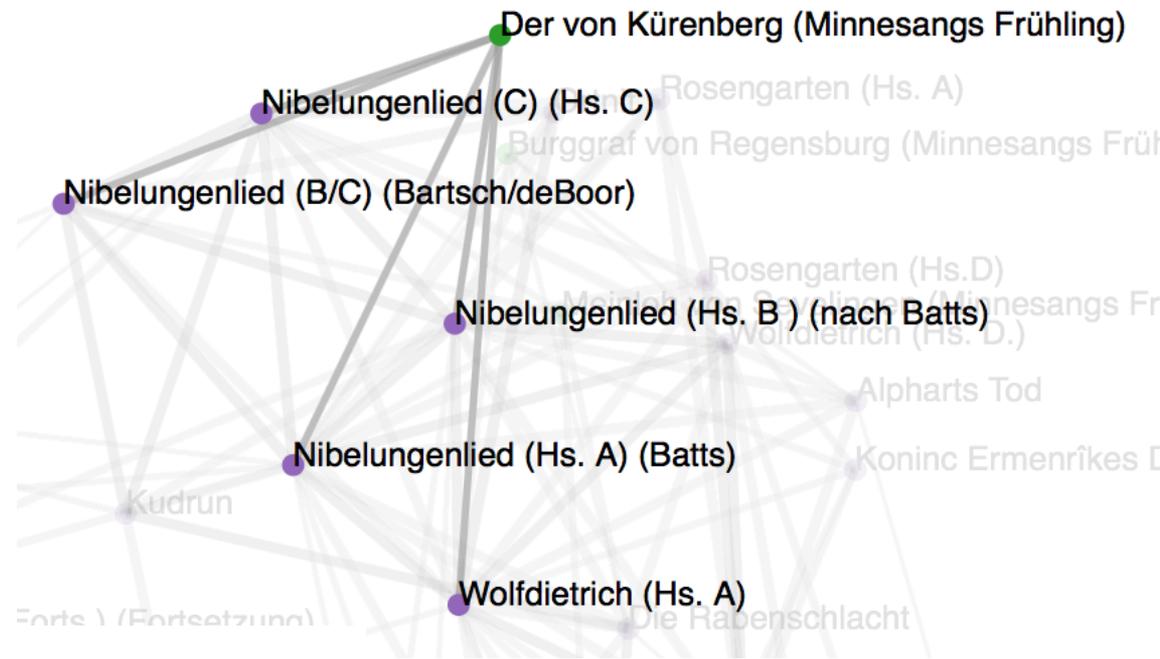
Table 4.2: Verse sequence frequency

# Cosinus-Maß auf TFIDF Werte

# Visualisierung

- <http://mhg-sound.appspot.com/>





sequence	<i>tfidf</i>
C-OOOO-C-C	0.091087
-C-C-OO-C-	0.089309
-C-OO-C-C-	0.085931
C-C-OOOO-C	0.085273
-C-C-XC-CO	0.077646

Table 4.4: Top five *tfidf* values for Der von Kurenberg

---

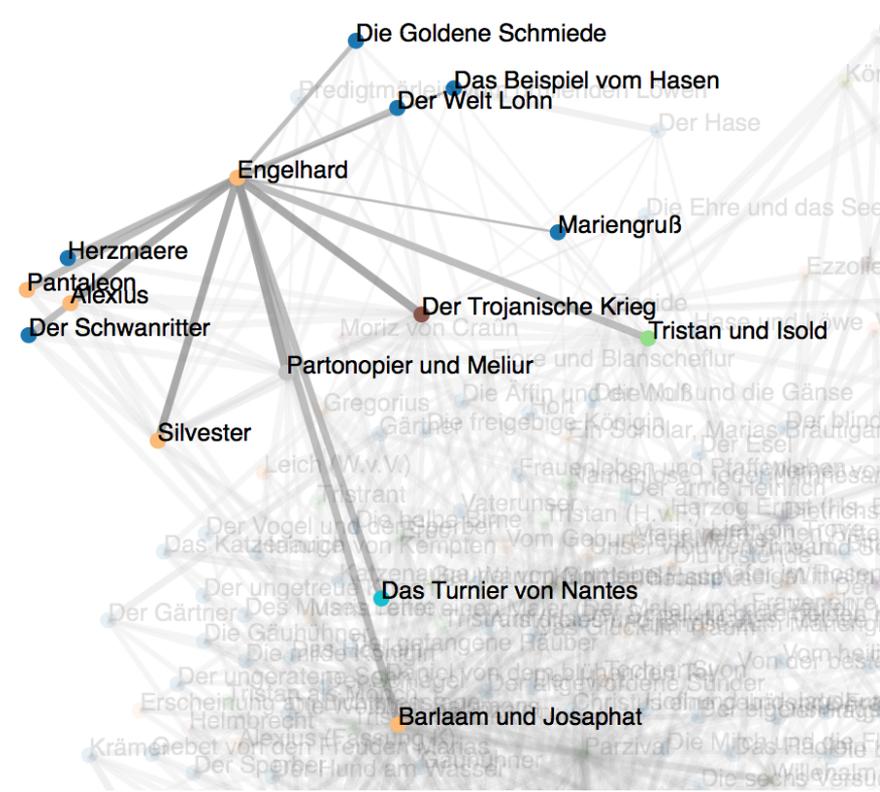
text	similarity to <i>NL</i> (B/C, Bartsch/deBoor)
<i>Nibelungenlied</i> C (Hennig)	.983
<i>Nibelungenlied</i> B (Batts)	.961
<i>Nibelungenlied</i> A (Batts)	.952
<i>Kudrun</i>	.879
<i>Wolfdietrich</i>	.877

---

Table 4.6: Five most similar texts to *Nibelungenlied* B/C, Bartsch/deBoor







sequence	<i>tfidf</i>
-C-XC-C-C-	0.108341
C-XC-C-C-C	0.107090
C-1C-C-C-C	0.099839
C-C-C-OC-C	0.090290
C-C-C-C-C-	0.084167

Table 4.10: Top five *tfidf* values for Konrad's *Engelhard*



# Interpretation

- chronologischer Aspekt
- formale Beziehungen
- Nachahmung

# Normales lexikalisches Verfahren

- Lemmata n-grams (unigrams, bigrams, trigrams)

# Korrelation

	unigrams	bigrams	trigrams
formal vs. lexical	.624 (.640)	.799 (.801)	.834 (.839)
top 20 overlap	21.8%	32.6%	36.2%

Table 4.25: Pearson's (and Spearman's)  $r$  for formal vs. lexical methods, and top 20 overlap

# Entfernung mittels Fuzzy-Matching

## > .85 Levenshtein-Distanz

Sequence: *CO-C-C-C-O*

*'herre ich hân von iu' ≈ 'herre ich hân von iu'*

Sequence: *-C-CO-C-XO*

*'wîp unde man â' ≈ 'wîp unde man ze''*

Sequence: *1C-C-C-OC-*

*'ob er noch rîters' ≈ 'ob der noch strîtes'*

# < .85 Levenshtein-Distanz

Sequence: *-XOC-C-C-O*

*'iuwer zuht und iu' ≠ 'biutet gein dem ei' ≠ 'sînen zoum nam si' (etc.)*

---

text	sim.
Rennewart	0.928854
Alexander (U.v.E) (Ulrich von Eschenbach)	0.923586
Wigalois, der Ritter mit dem Rade	0.923414
Parzival	0.916120
Daniel von dem blühenden Tal	0.915249
Barlaam und Josaphat	0.914278
Lanzelet	0.913639
Karl der Grosse	0.913152
Erec	0.908859
Willehalm (Wolfram)	0.905695

---

---

sequence	sim.
Parzival	0.771731
Willehalm (Wolfram)	0.766083
Tristan und Isold	0.763064
Alexander (U.v.E) (Ulrich von Eschenbach)	0.758367
Erec	0.756541
Der Renner	0.753834
Biterolf und Dietleib	0.753810
Barlaam und Josaphat	0.753227
Wigalois, der Ritter mit dem Rade	0.752769
Wilhelm von Wenden	0.751707

---

Table 4.21: Top ten cosine similarities after removal of top ten

	$r$	overlap 10	overlap 20
<i>Iwein</i>	.999 (.999)	100%	85%
<i>Parzival</i>	.999 (.999)	100%	85%
Top 10	.991 (.991)	60%	70%

Table 4.22: Pearson's (and Spearman's)  $r$  for cosine similarity, and top 20 overlap before and after for *Erec*

text	overlap
Eneide	0.80
Herzmaere	0.80
Tristan und Isold	0.70
Gregorius	0.70
Willehalm (Wolfram)	0.70
Erec	0.65
Karl der Grosse	0.65
Das Katzenauge	0.65
Tochter Syon	0.65
Der Trojanische Krieg	0.65
Reinfried von Braunschweig	0.60
Willehalm (U.v.T.)	0.60
Walther von Klingen	0.60
Der arme Heinrich	0.60
Flore und Blanscheflur	0.60
Tristrant	0.60
Engelhard	0.60
...	...
Boppe	0.10
Eckenlied (Fassung L.)	0.10
Hugo von Mülndorf	0.05
Burggraf von Regensburg (Minnesangs Frühling)	0.05
Herzog Ernst (Hs.D, strophig)	0.05
Der Jüngere Titurel	0.05
Der Wunderer	0.05
Dietmar der Setzer	0.05
Titurel	0.05
Lanzilet	0.05
Meinloh von Sevelingen (Minnesangs Frühling)	0.05
Virginal	0.05
Der von Wengen	0.05
Lohengrin	0.00
Die Klage der Kunst	0.00
Der Ritterspiegel	0.00

Table 1: Top 20 overlap of formal and lexical methods

# Formeln

sequence	<i># removed</i>
-1C-C-C-C-	354
-C-1C-C-C-	196
-XC-C-C-C-	195
-C-XC-C-C-	172
-C-OOC-1C-	164
C-C-C-C-C-	150
-OOC-1C-C-	119
-C-C-C-CC-	111
-C-C-C-OC-	96
-C-C-OOC-1	92

Table 4.24: Most frequently removed sequences from top 10 similarity

sequence	<i># removed</i>
der künec artûs	76
dô sprach der künec	28
sî sprach lieber herre	24
als ich iu sage	24
ich ie wart geborn	22
lîbe und mit guote	20
dô sprach der herre	20
ich weiz wol daz	19
nâch âventiure reit	18
biderbe unde guot	18
unz an mînen tôt	16
dô sprach diu künegîn	16
bete und mîn gebot	16
hân sô wil ich	15
den künec artûs	15
sprach diu künegîn	14
mit worten und mit	14
swaz ir gebietet daz	14
swaere daz er im	14
in willekomen sîn	14
wil ich iu sagen	14
im saelde und êre	13
in dem lande nie	12
nâch rîterlîchen siten	12
der künec artûs ze	12
an dem lîbe und	12
dô sprach der rîter	12
einen slac daz er	12
waere geschehen ob	12
ich iu wil sagen	12

Table 1: Most frequently removed matches from top 10 similarity