

Klanglandschaften im Minnesang: Silbentrennung als Werkzeug zur Literaturanalyse

Christopher Hench

UC Berkeley, Department of German

Motivation

Wie können Mediävisten Computer-gestützte Textanalyse ausführen?

Wie können Mediävisten Klang forschen?

Die Silbe

Warum die Silbe?

Walther von der Vogelweide (E Wa 78)

Frouwe, ir lâ̄t iuch nit verdriezen
mîne rede, ob sie gefüege sî.
möht is wider iuch geniezen,
sô wære ich den besten gerne bî.
wizzet, daz ir schœne sît.
habt ir, als ich mich verwæne,
güete bî der wolgetæne?
waz denne an ir einer êren lît!

Frauwe ir lat [˙]vch nit [˙]v dri[˙]zzē
mine rede· ob fie gefüge fin·
möht iz wider [˙]vch geni[˙]zzē·
fo were ich den beften gerne bi·
wizzet daz ir fchône fit·
habt ir als ich mich [˙]v wene·
güte bi der wolgetene·
waz denne an ir ein[˙] eren lit·

46 word(s)
70 syllable(s)
1.52173913043 syllable(s) per word

17 light syllable(s)
53 heavy syllable(s)
24.2857142857 percent light

34 open syllable(s)
36 closed syllable(s)
48.5714285714 percent open

46 word(s)
70 syllable(s)
1.52173913043 syllable(s) per word

25 light syllable(s)
45 heavy syllable(s)
35.7142857143 percent light

31 open syllable(s)
39 closed syllable(s)
44.2857142857 percent open

A Wa 13 = L 69,1

IV

Saget mir ieman wc ift / minne

fo wift ich gerne me·

fwer sich rehte nv verfinne·

d^r berihte mich wie / t^v t fi fo we·

5 minne ift minne t^v t fi wol·

t^v t fi we fo enheizet fi niht rehte / [...]ne

foz enweiz ich wie fi danne heizen fol·

29 offene Silben

35 geschlossene Silben

45.31% offen

C Wa 245 (241) = L 69,1

Zitier

II

Saget mir ieman wc ift mīne·

fo weft / ich gerne o^v ch darvmbe me·

fwer sich / rehte nv v^rfinne·

der berihte rehte mich / wie t^u t fi we·

5 mīne ift mīne t^u t fi wol· /

t^u t fi we fone heiffet fi niht mīne fvs / enweis ich ~~we~~ wie fi danne heiffē

fol· /

A 13

S 14rb,5

31 offene Silben

38 geschlossene Silben

44.93% offen

S 14rb,5 = L 69,1

¶ Saget mir yemā waz ift mīne // [14va]
Wijft ichs ein deil fo wift ichs g⁷ne me| /
Der sich bas dan ich v⁷inermee /
Berichte mich dorch was zi doyt zo we /
5 Mīne ift mīne deyt fi wol /
Deyt zi wee zone heifet n3 rechte mīne
Suf in weys ich n3 wie fi heyfē fol

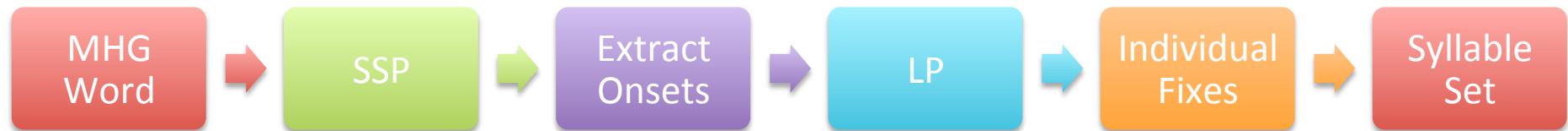
26 offene Silben
42 geschlossene Silben
38.24% offen

Methode

- Silbentrennung
 - Phonotaktik
 - Ästhetik
 - Klanglandschaften
- Skandierung
 - automatische Skandierung
 - Stylometrie

Silbentrennung

- Sonority Sequencing Principle (SSP)
- Legality Principle (LP)



Die mittelhochdeutsche Silbentrennung

be-gun-de, senf-ter, um-be, ker-zen, ge-dul-de-keit
sonority sequencing principle

ta-ge, zwî-vel, sê-le, ge-nant, sa-gen, â-ne
onset maximization

velare Nasale

lange

NHD [lan̩ə] MHD [lan̩gə]

heterosyllabisch im MHD: *lan-ge*

ambisyllabischer velarer Nasal im NHD

Geminaten

vallen/fallen: NHD [falən] MHD [falːən]

MHG *vallen* kann als *val-len* getrennt werden, im NHD mit einem ambisyllabischen Konsonant

Geminat oder ambisyllabisch?

Fragen: Affrikaten

1.[zɪ.tsən] *sitzen*

Vorteil: eine monophonemische Affrikate folgt dem Onset-
Maximierungsprinzip

Nachteil: ergibt NHD [zɪ:tsən]

361 si sêre solde letzen
 X / X' X / X' X / ---' / X`
362 und vil gar ent - setzen
 X' X / X' X / ---' / X`

Der arme Heinrich

Fragen: Affrikaten

2. [zɪts.ən] monophonemisch, der linken Silbe gehörend, löst das Problem der offenen Silbe, aber stößt gegen das Onset-Maximierungsprinzip

cf. *le-ben*

Fragen: Affrikaten

3. /zɪtsən/ monophonemisch, ambisyllabisch
4. /zɪt.sən/ biphonemische, heterosyllabische Phoneme
– ergibt eine geschlossene erste Silbe und eine zweite Silbe mit einem legalen Onset

Für den Algorithmus gibt es *keinen Unterschied*:
sit-zen kann entweder ambisyllabisch oder mit zwei heterosyllabischen Phonemen angesehen werden

Fragen: s + t

- beste* → *be-ste* oder *bes-te*?
- be-ste* ergibt NHD [beː.stə]
- metrischer Beweis unterstützt *bes-te*:

- 97 Dirre werlde veste,
 X' X / X' X / ---' / X`
- 98 ir stæte und ir beste
 X/ X' X / X' X / ---' / X`

Der arme Heinrich

Morphemgrenzen und illegale Onsets

- geste* → *ges-te* aber *gestalt* → *ge-stalt*
- wîplich* → *wîp-lich*, nicht *wî-plich*
- lîch* und *-heit* (einzeln behandelt)
- enweiz* (*ich enweiz*) → *en-weiz*, nicht *e-nweiz*

Genauigkeit

99.4%

95% Konfidenzintervall 98.9% to 99.9%

Following onsets > .02 percent deemed 'legal':

n - sl - s - br - tr - str - q - ph - w - h - gl - schr - b - g - m - pr - fr - kn - fl - r - p - spr - d - vr - vl -
sch - zw - kr - t - sm - sn - c - l - kl - pl - bl - tw - f - v - k - pfl - z - dr - st - sw - cl - sp - j - gr - pf -
phl - tj

ist zwî-vel her-zen nâch-ge-bûr daz muoz der sê-le wer-den sûr ge-smæ-het un-de ge-zi-e-ret ist swâ sich par-ri-e-ret
un-ver-za-get man-nes muot als a-gel-stern var-we tuot der mac den-noch we-sen geil wand an im sint be-i-diu teil des
hi-mels und der hel-le der un-stæ-te ge-sel-le hât die swar-zen var-we gar und wirt och nâch der vin-ster var sô ha-bet
sich an die blan-ken der mit stæ-ten ge-dan-ken diz vli-e-gen-de bî-spel ist tum-ben li-u-ten gar ze snel si-ne mu-gens
niht er-den-ken wand ez kan vor in wen-ken reh-te al-sam ein schel-lec ha-se zin an-der-halp a-me gla-se ge-lî-chet und
des blin-den troum die ge-bent ant-lût-zes roum doch mac mit stæ-te niht ge-sîn dir-re trû-e-be lîh-te schîn er ma-chet
kur-ze frô-u-de al-wâr wer ro-u-fet mich dâ nie kein hâr ge-wuohs in-ne an mî-ner hant der hât vil nâ-he grif-fe er-
kant sprich ich gein den vorh-ten och daz glî-chet mî-ner wit-ze doch wil ich tri-we vin-den al-dâ si kan ver-swin-den
als viur in dem brun-nen unt daz tou von der sun-nen auch er-kan-te ich nie sâ wî-ger man er-ge-mâh-te ger-ne kün-de hân

Following onsets > .02 percent deemed 'legal':

sq - p - str - sc - s - v - fl - gl - h - br - d - t - f - dr - thr - r - n - rh - gr - c - m - g - ch - tr - z - sp -
spr - st - ph - cl - q - b - cr - bl - fr - th - phr - pr - l - pl

in no-va fert a-ni-mus mu-ta-tas di-ce-re for-mas cor-po-ra di co-ep-tis nam vos mu-ta-stis et il-las ad-spi-ra-te meis
pri-ma-que ab o-ri-gi-ne mun-di ad mea per-pe-tuum de-du-ci-te tem-po-ra car-men an-te ma-re et ter-ras et quod te-git
om-nia ca-e-lum u-nus e-rat toto na-tu-rae vul-tus in or-be quem dix-e-re chaos ru-dis in-di-ge-sta-que mo-les nec qu-
ic-quam ni-si pon-dus i-ners con-ge-sta-que e-o-dem non be-ne i-unc-ta-rum di-scor-dia se-mi-na re-rum nul-lus ad-huc
mun-do pra-e-be-bat lu-mi-na ti-tan nec no-va cre-scen-do re-pa-ra-bat cor-nua pho-e-be nec cir-cum-fu-so pen-de-bat in
a-e-re tel-lus pon-de-ri-bus li-bra-ta suis nec brac-chia lon-go mar-gi-ne ter-ra-rum por-rex-e-rat am-phi-tri-te ut-
que e-rat et tel-lus il-lis et non-tus et aer sic e-rat in-sta-bi-lis tel-lus in-na-bi-lis un-da lucis e-gens aer nul-

Following onsets > .02 percent deemed 'legal':

dw - fr - pr - kr - pl - g - b - j - sn - q - wl - spr - xr - gr - pw - sl - tr - hv - dr - d - l - swn - wr - sm - phm
- n - br - hl - sk - t - z - p - php - p - w - bl - mk - lk - s - k - sp - sw - r - tw - f - jn - st - mt - hr - hn -
th - h - c - pr - m

ak a-na lu-kar-na-sta-pin jah li-u-teip al-laim þaim in þam-ma gar-da swa li-uht-jai li-u-hap iz-war in an-dwa-irp-ja
man-ne ei ga-sa-ihv-a-i-na iz-wa-ra go-da wa-urs-twa jah ha-uh-ja-i-na at-tan iz-wa-ra-na þa-na in hi-mi-nam ni hug-
jaip ei qem-jau ga-ta-i-ran wi-top a-ip-þau pra-u-fe-tuns ni qam ga-ta-i-ran ak us-full-jan a-men auk qi-þa iz-wis und
þa-tei u-sle-i-þip hi-mins jah a-ir-þa jo-ta ains a-ip-þau ains striks ni u-sle-i-þip af wi-to-da un-te al-la-ta wa-ir-
þip ip saei nu ga-ta-i-rip a-i-na a-na-bu-sne þi-zo min-ni-sto-no jah la-is-jai swa mans min-ni-sta ha-i-ta-da in þi-u-
dan-gard-jai hi-mi-ne ip saei ta-u-jip jah la-is-jai swa sah mi-kils ha-i-ta-da in þi-u-dan-gard-jai hi-mi-ne qi-þa auk
iz-wis þa-tei ni-bai ma-na-gi-zo wa-ir-þip iz-wa-ra-i-zos ga-ra-ih-teins þau þi-ze bo-kar-je jah fa-re-i-saie ni þau
qi-mib in hi-u-dan-gard-jai hi-mi-ne þa-u-si-de-dub þa-tei qi-þan ist þaim a-i-ri-zam ni ma-urbr-jais ih saei ma-ur-

Following onsets > .02 percent deemed 'legal':

cl - m - f - iv - t - j - tr - g - h - ll - i - b - pl - c - p - l - n - ch - fr - cr - fl - s - gl - d - gr - r - br -
q - bl - pr - z - v

pró-lo-go de-so-cu-pa-do lec-tor sin ju-ra-men-to me pod-rás creer que qu-i-si-e-ra que es-te li-bro co-mo hi-jo del
en-ten-di-mi-en-to fu-e-ra el más her-mo-so el más ga-llar-do y más dis-cre-to que pu-di-e-ra i-ma-gi-nar-se pe-ro no
he po-di-do yo con-tra-ve-nir al or-den de na-tu-ra-le-za que en e-lla ca-da co-sa en-gend-ra su se-me-jan-te y a-sí
qué pod-rá en-gend-rar el es-té-ri-l y mal cul-ti-va-do in-ge-nio mío si-no la his-to-ria de un hi-jo se-co a-ve-lla-na-
do an-to-ja-di-zo y lle-no de pen-sa-mi-en-tos va-rios y nun-ca i-ma-gi-na-dos de o-tro al-gu-no bien co-mo quien se
en-gend-ró en u-na cár-cel don-de to-da in-co-mo-di-dad ti-e-ne su a-si-en-to y don-de to-do tris-te ru-i-do ha-ce su
ha-bi-ta-ción el so-si-e-go el lu-gar a-pa-ci-ble la a-me-ni-dad de los cam-pos la se-re-ni-dad de los ci-e-los el mur-
mu-rar de las fu-en-tes la qu-i-e-tud del es-pí-ri-tu son gran-de par-te pa-ra que las mu-sas más es-té-ri-les se mu-
es-tren fe-cun-das y o-frez-can par-tos al mun-do que le col-men de ma-ra-vi-lla y de con-ten-to a-con-te-ce te-ner un
pad-re un hi-jo feo y sin gra-cia al-gu-na y el a-mor que le ti-e-ne le po-ne u-na ven-da en los o-jos pa-ra que no vea
sus fal-tas an-tes las juz-ga por dis-cre-ci-o-nes y lin-de-zas y las cu-en-ta a sus a-mi-gos por a-gu-de-zas y do-na-
i-res pe-ro yo que a-un-que pa-rez-co pad-re soy pad-ras-tro de don qu-i-jo-te no qu-i-e-ro ir-me con la cor-ri-en-te
del u-so ni su-pli-car-te ca-si con las lá-ari-mas en los o-ios co-mo o-tros ha-cen lec-tor ca-rí-si-mo que per-do-nes

Following onsets > .02 percent deemed 'legal':

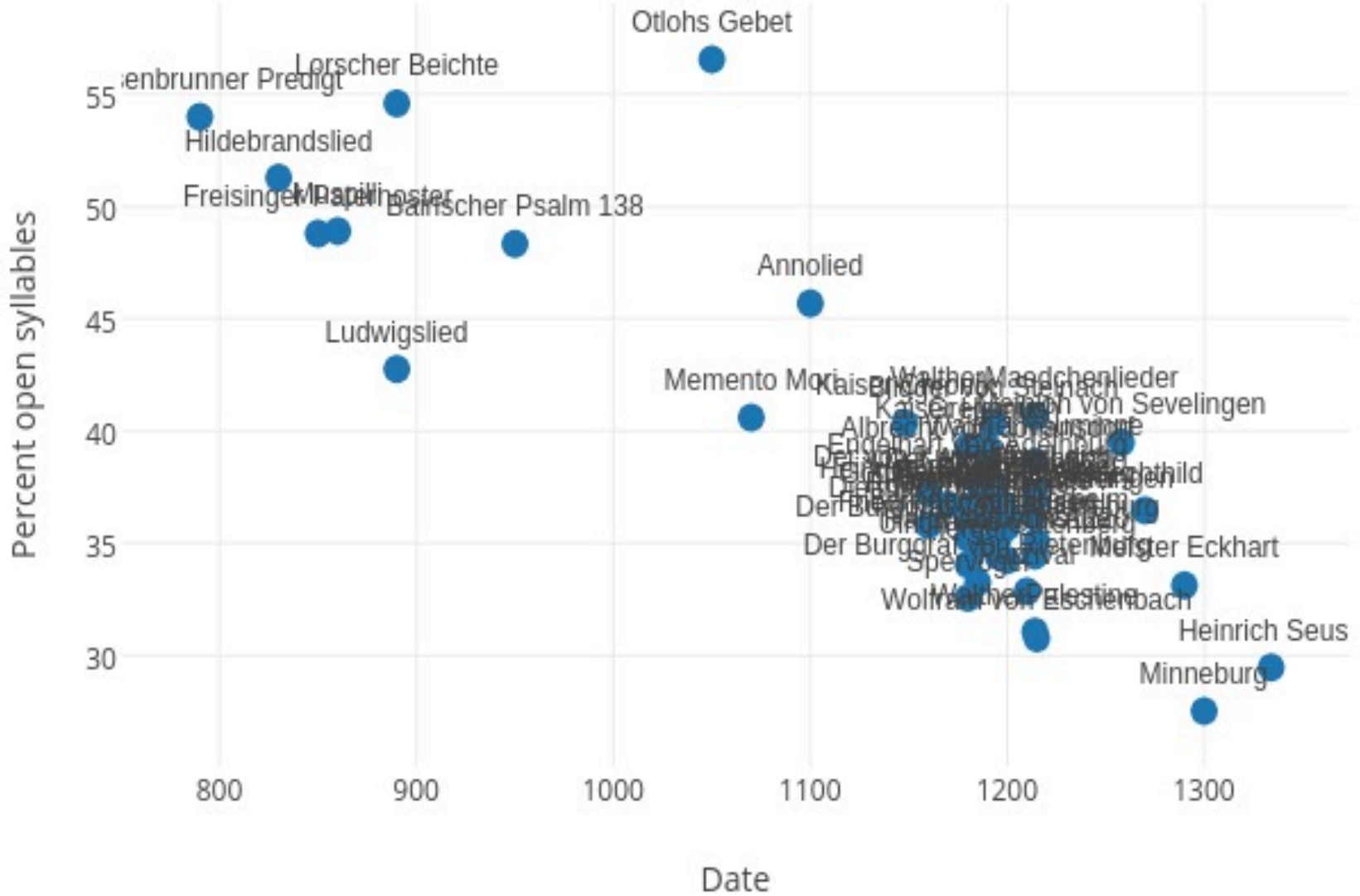
spr - kv - c - lj - sc - vr - tv - r - g - n - sm - sv - skj - k - br - kn - sj - pl - sch - dj - fl - tr - sn - h -
skr - b - hskr - dr - nj - kr - str - hm - t - sk - x - j - rdr - sl - gl - gn - tj - gr - q - m - v - d - p - bj - gj
- kl - stj - hj - fj - mj - pr - s - f - fr - l - st - bl - sp

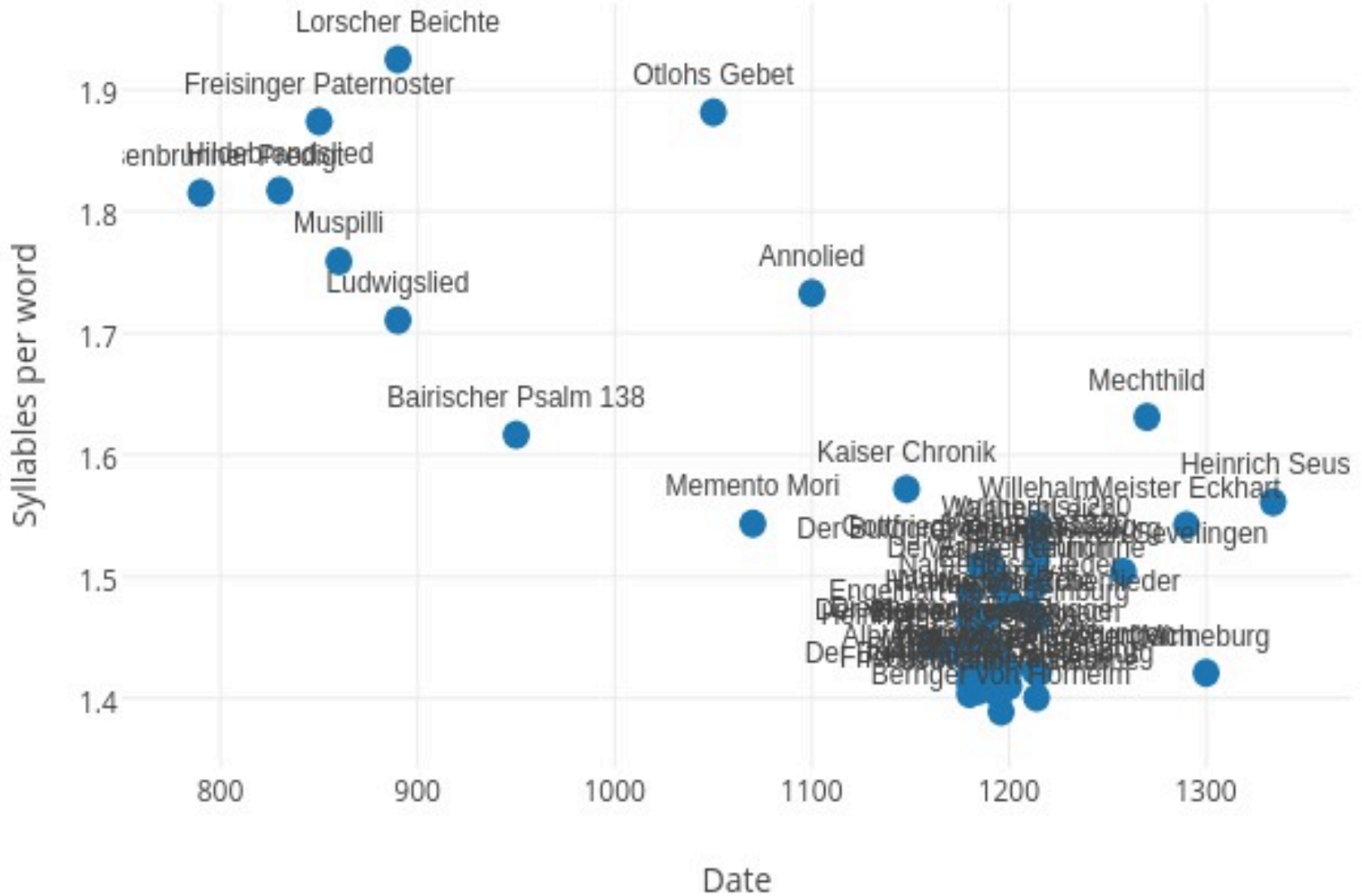
för-sta ka-pit-let stock-holm i få-gel-per-spek-tiv det var en af-ton i bör-jan av maj den lil-la träd går-den på mo-
se-bac-ke ha-de än-nu ic-ke bli-vit öpp-nad för all-män-he-ten och ra-bat-ter-na vo-ro ej upp-gräv-da snö drop-par-na
ha-de ar-be-tat sig upp ge-nom fjo-lå-rets löv sa-mun-gar och höl-lo just på att slu-ta sin kor-ta verk-sam het för att
läm-na plats åt de öm-tå-li-ga-re saf-frans blom-mor-na vil-ka ta-git skydd un-der ett o-frukt-samt pä-ron-träd sy-re-
ner-na vän-ta-de på syd-lig vind för att få gå i blom men lin-dar-na bjö-do än-nu kär-leks-fil-ter i si-na o-brust-na
knop-par åt bo-fin-kar-na som bör-jat byg-ga si-na lav-kläd-da bon mel-lan stam och gren än-nu ha-de in-gen män-sko-fot
tram-nat sand-nån-gar-na se-dan si-sta vin-terns snö nått hort och där-för lev-des ett o-he-svä-rat liv dä-rin-ne av

Following onsets > .02 percent deemed 'legal':

ch - ph - thr - g - gl - l - cr - n - s - spr - dw - sq - th - t - tw - st - sc - tr - mr - scr - v - pl - j - kn - sn
- q - c - dr - sw - gr - fl - b - w - wr - p - br - cl - r - fr - h - spl - k - pr - bl - m - f - sp - sh - wh - sk -
sm - sch - str - d - sl

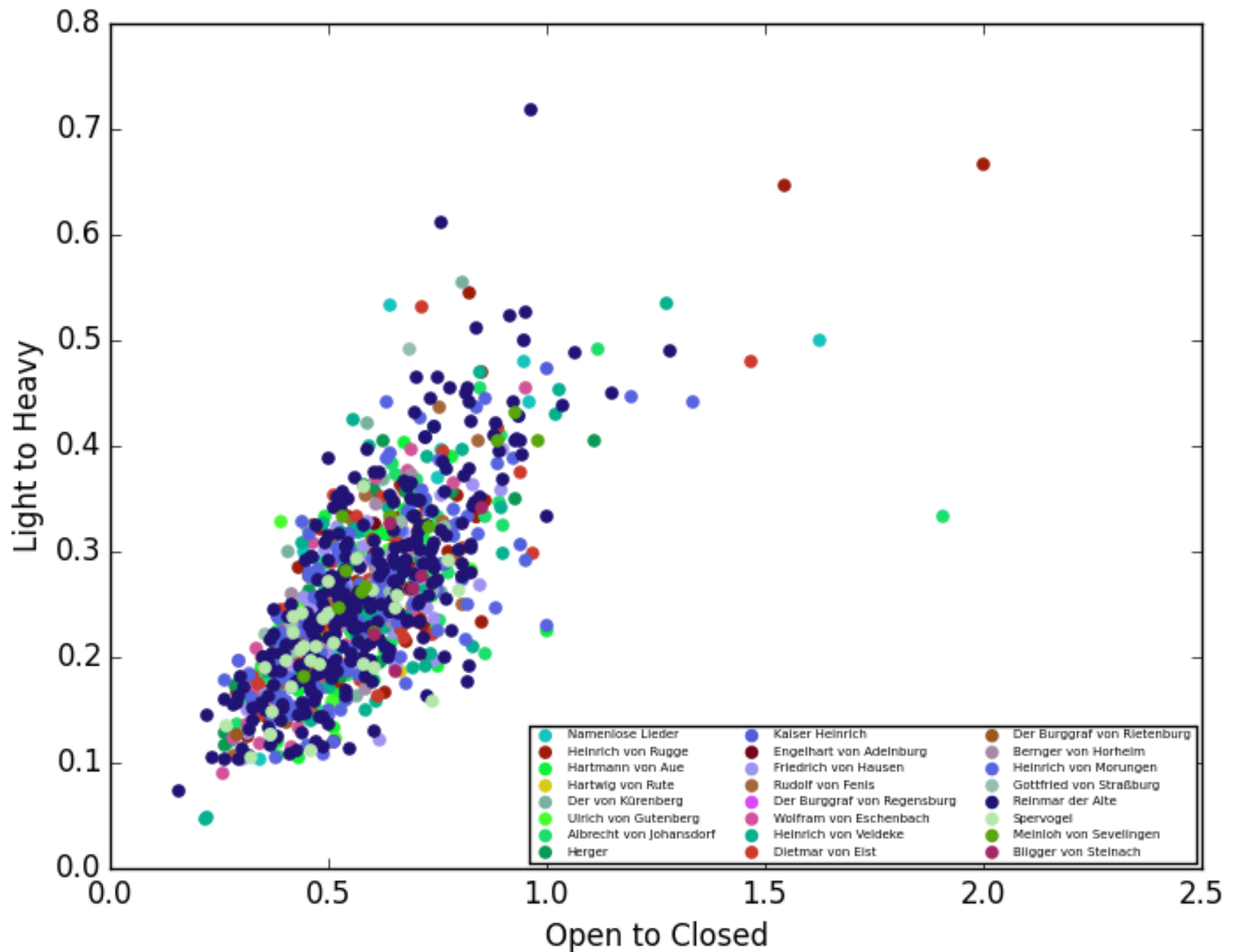
my fa-ther had a small e-sta-te in not-ting-ham-shi-re i was the third of fi-ve sons he sent me to em-ma-nuel col-le-ge
in cam-brid-ge at fo-ur-teen years old whe-re i re-si-ded three years and ap-plied my-self clo-se to my stu-dies but
the char-ge of ma-in-ta-i-ning me al-though i had a ve-ry scan-ty al-lo-wan-ce being too great for a nar-row for-tu-ne
i was bound ap-pren-ti-ce to ja-mes ba-tes an e-mi-nent sur-geon in lon-don with whom i con-ti-nued four years and my
fa-ther now and then sen-ding me small sums of mo-ney i laid them out in le-ar-ning na-vi-ga-tion and o-ther parts of
the ma-the-ma-tics u-se-ful to tho-se who in-tend to tra-vel as i al-ways be-li-e-ved it would be so-me ti-me or o-ther





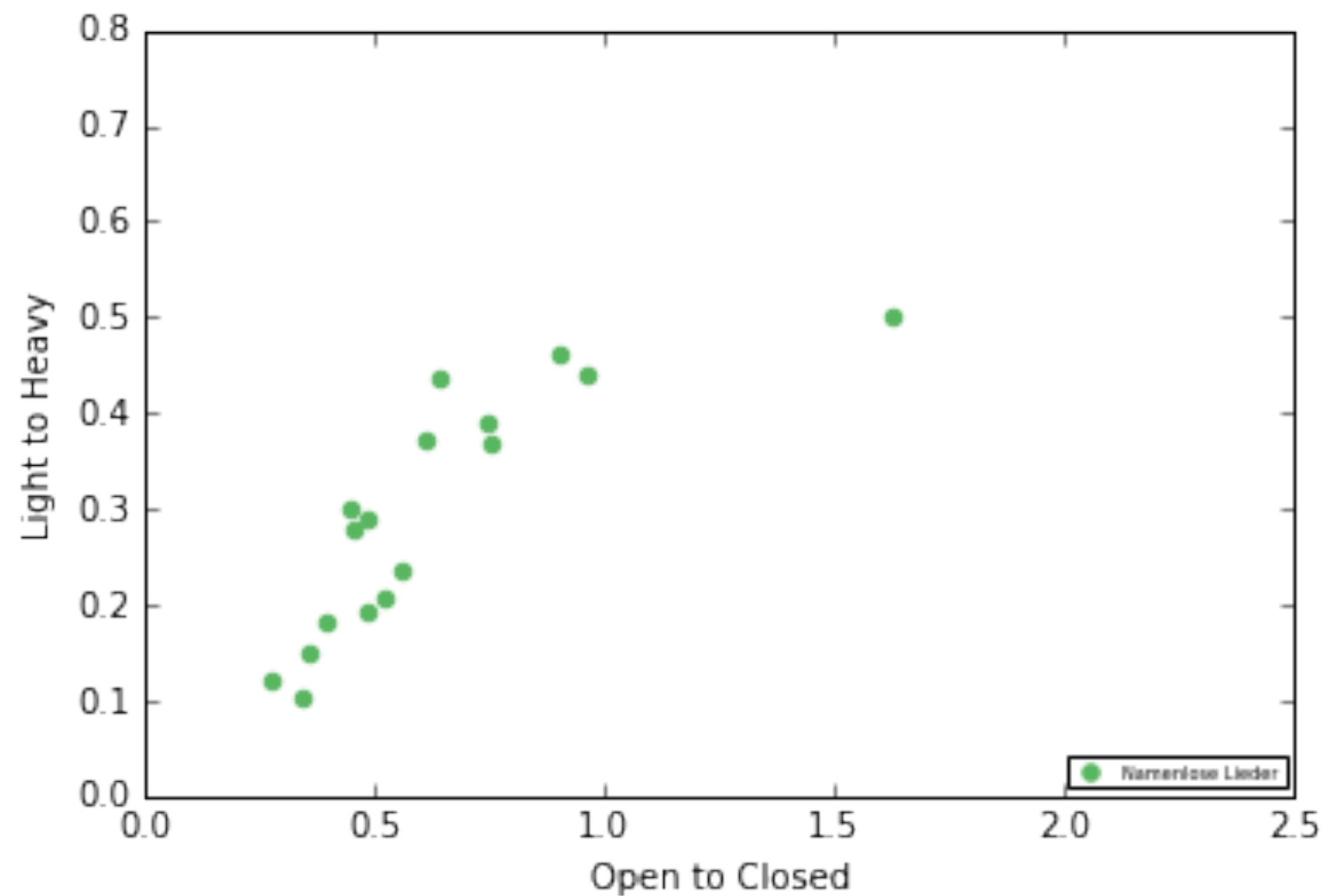
Des Minnesangs Frühling

The Sound of Minnesang

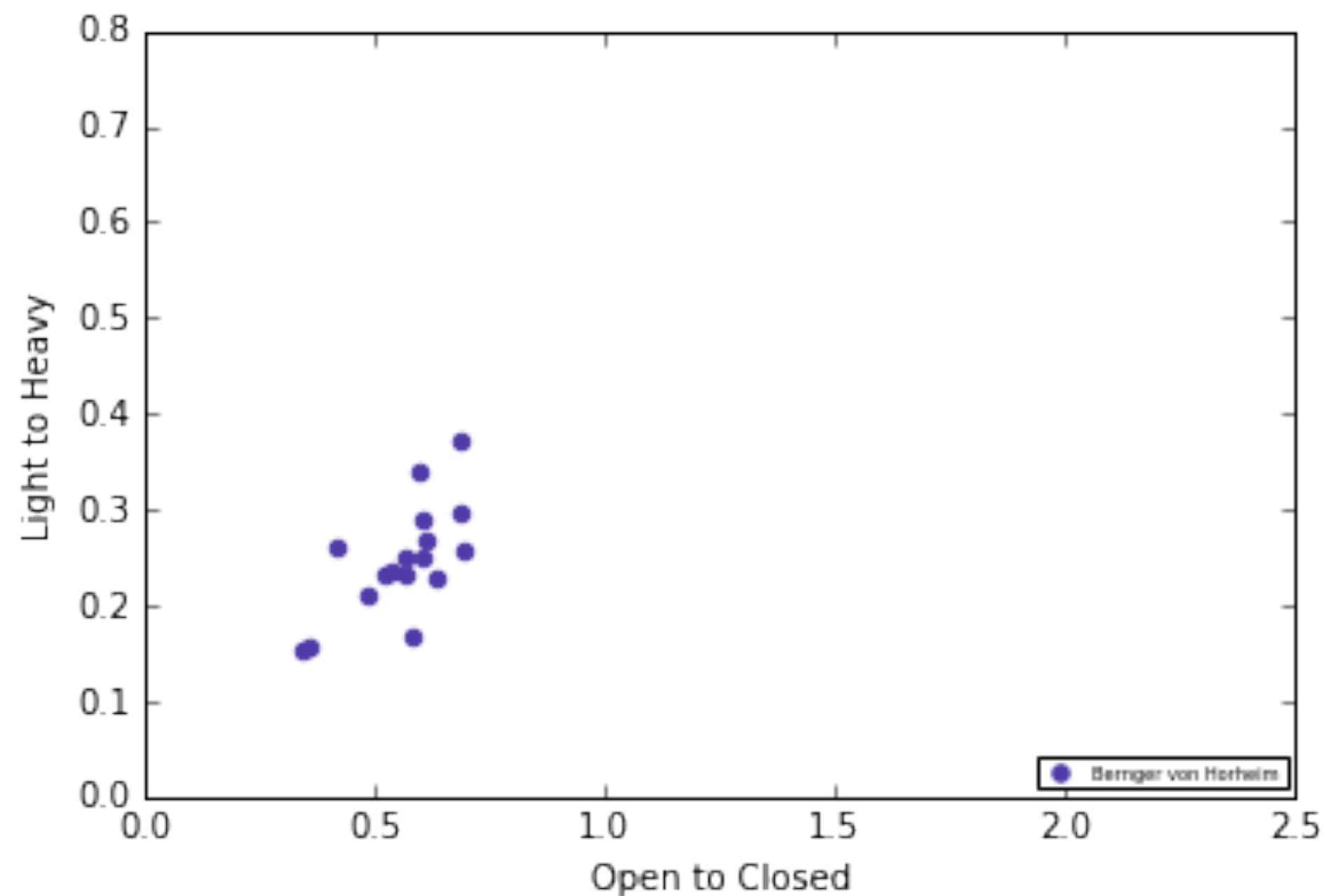


	std.
Namenlose Lieder	0.32252297
Albrecht von Johansdorf	0.26094126
Heinrich von Ruge	0.24749699
Kaiser Heinrich	0.21111849
Dietmar von Eist	0.20239976
Herger	0.18801148
Heinrich von Morungen	0.18489413
Heinrich von Veldeke	0.18388716
Meinloh von Sevelingen	0.17664027
Reinmar der Alte	0.17271587
Der Burggraf von Regensburg	0.17085937
Gottfried von Straßburg	0.16177029
Wolfram von Eschenbach	0.15703249
Friedrich von Hausen	0.15549111
Hartmann von Aue	0.14473299
Hartwig von Rute	0.14098614
Der von Kürenberg	0.13423448
Rudolf von Fenis	0.13342579
Spervogel	0.13038709
Der Burggraf von Rietenburg	0.1286002
Ulrich von Gutenberg	0.11533118
Bernger von Horheim	0.10646698
Bligger von Steinach	0.08030074
Engelhart von Adelnburg	0.07857598

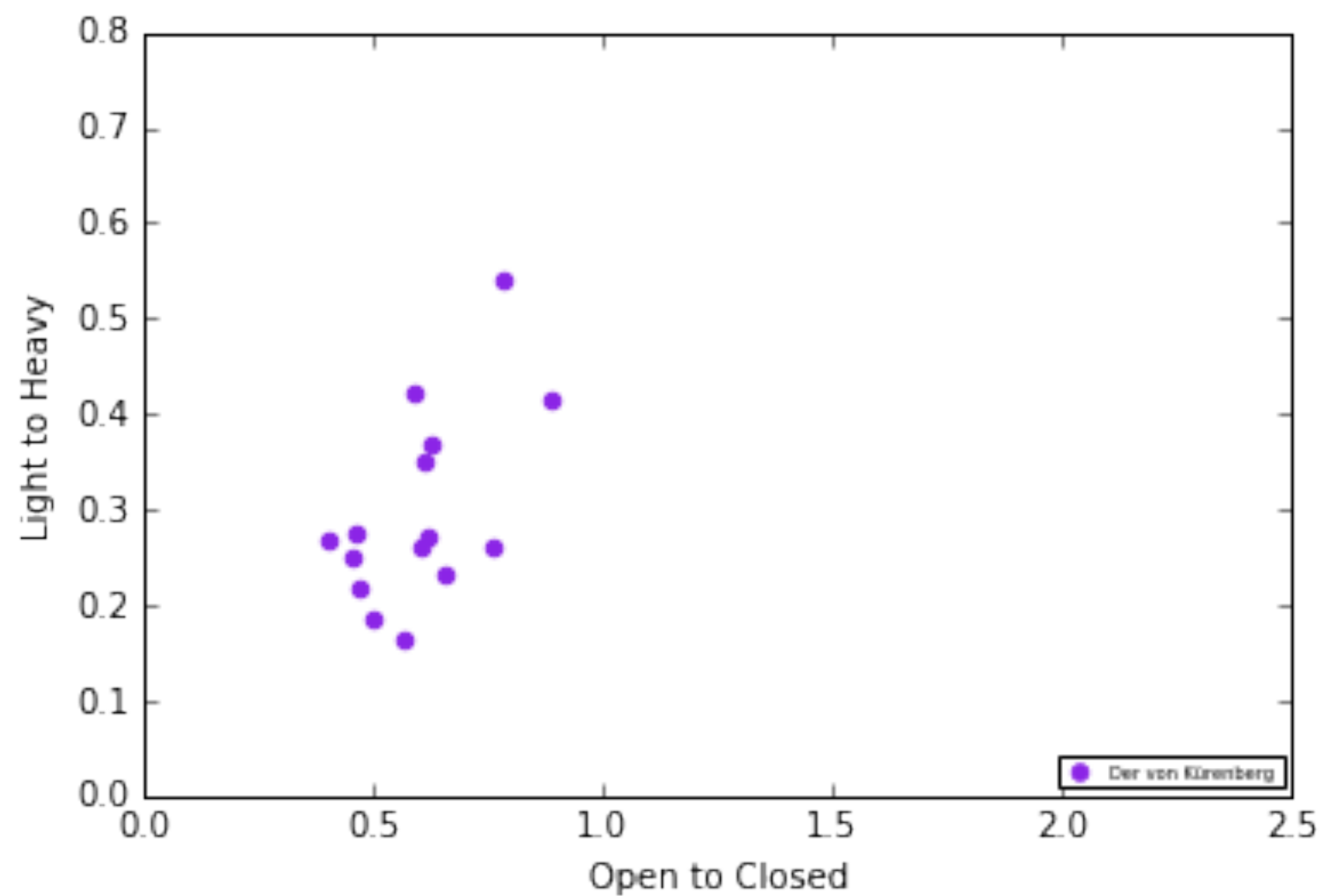
The Sound of Minnesang



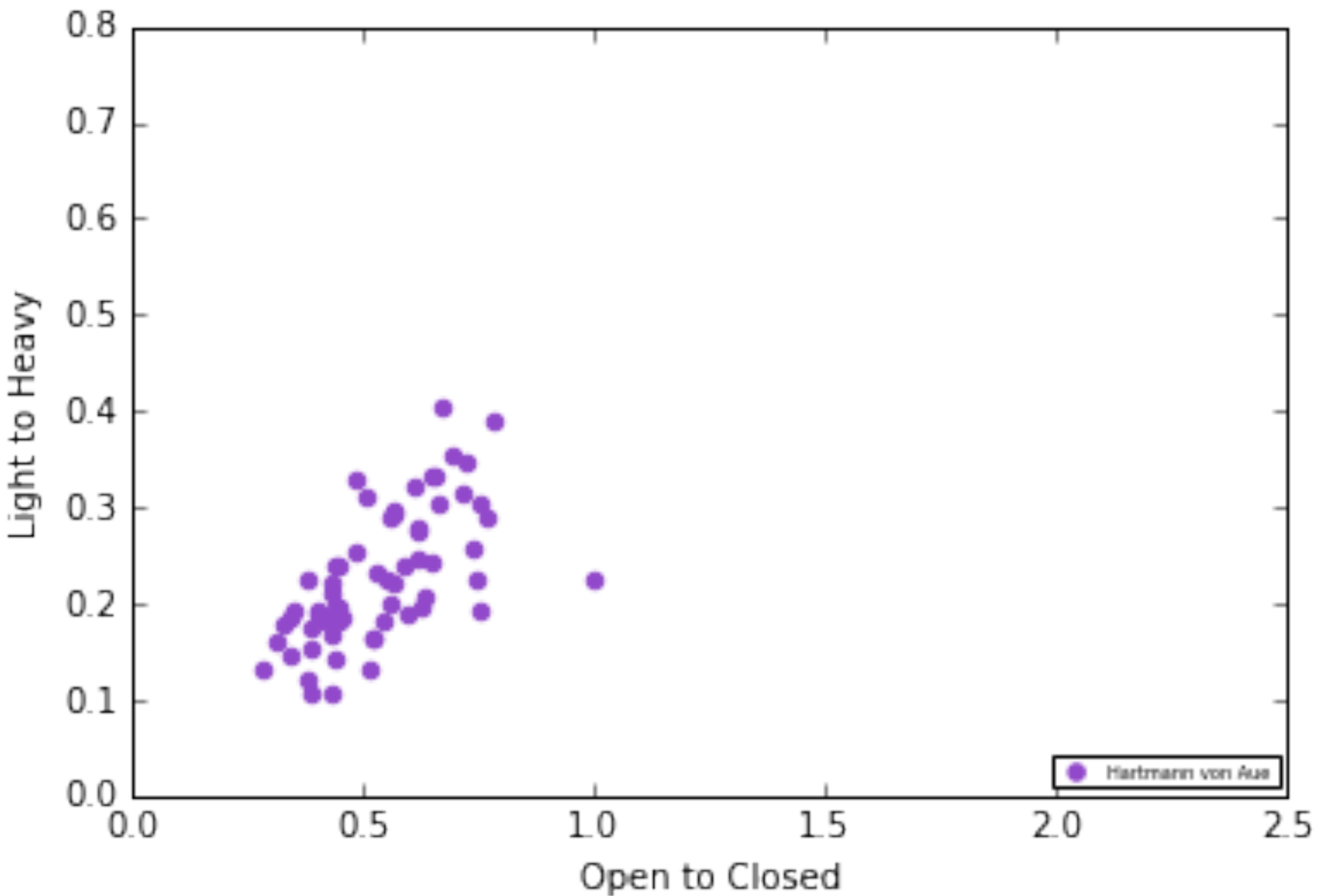
The Sound of Minnesang



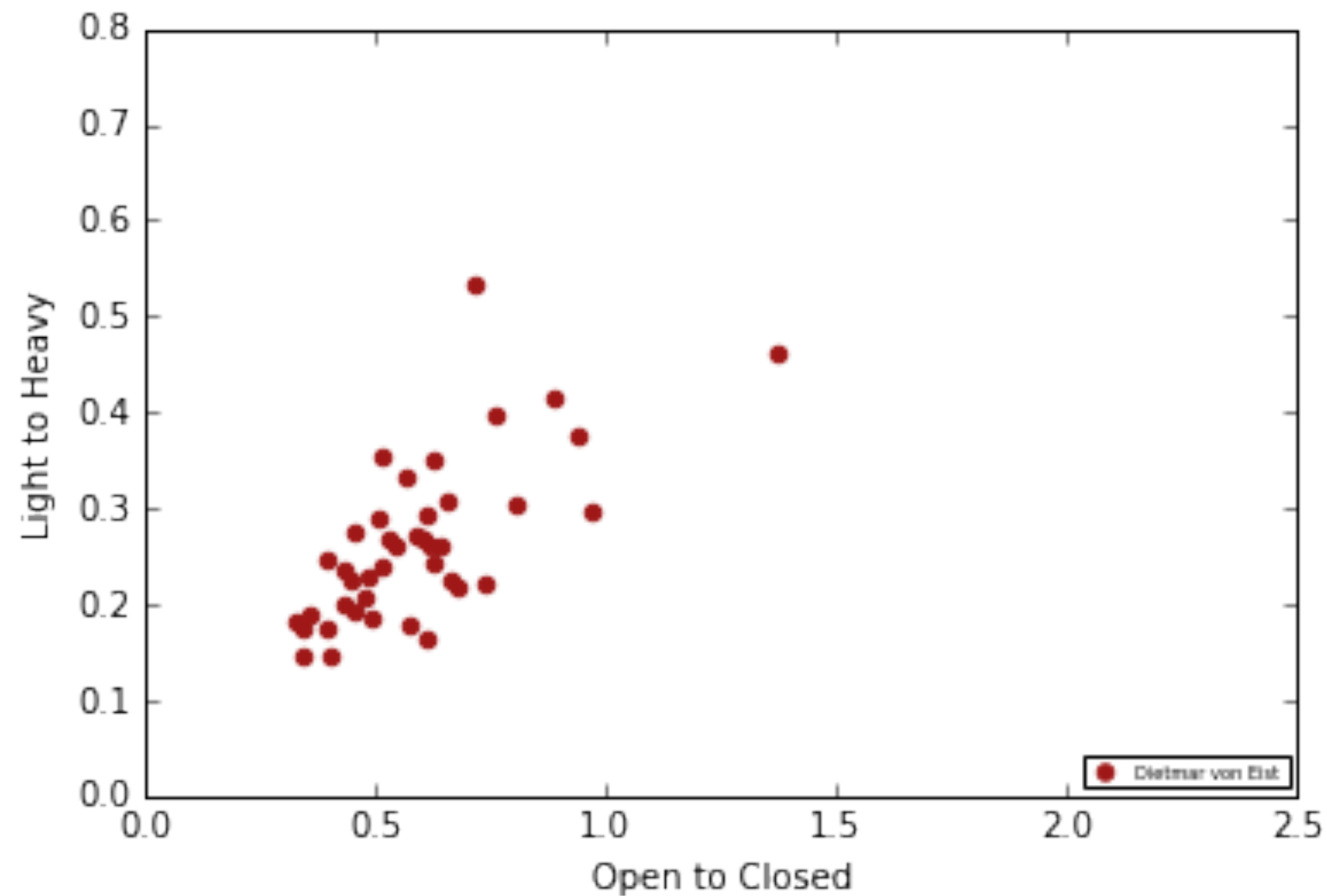
The Sound of Minnesang



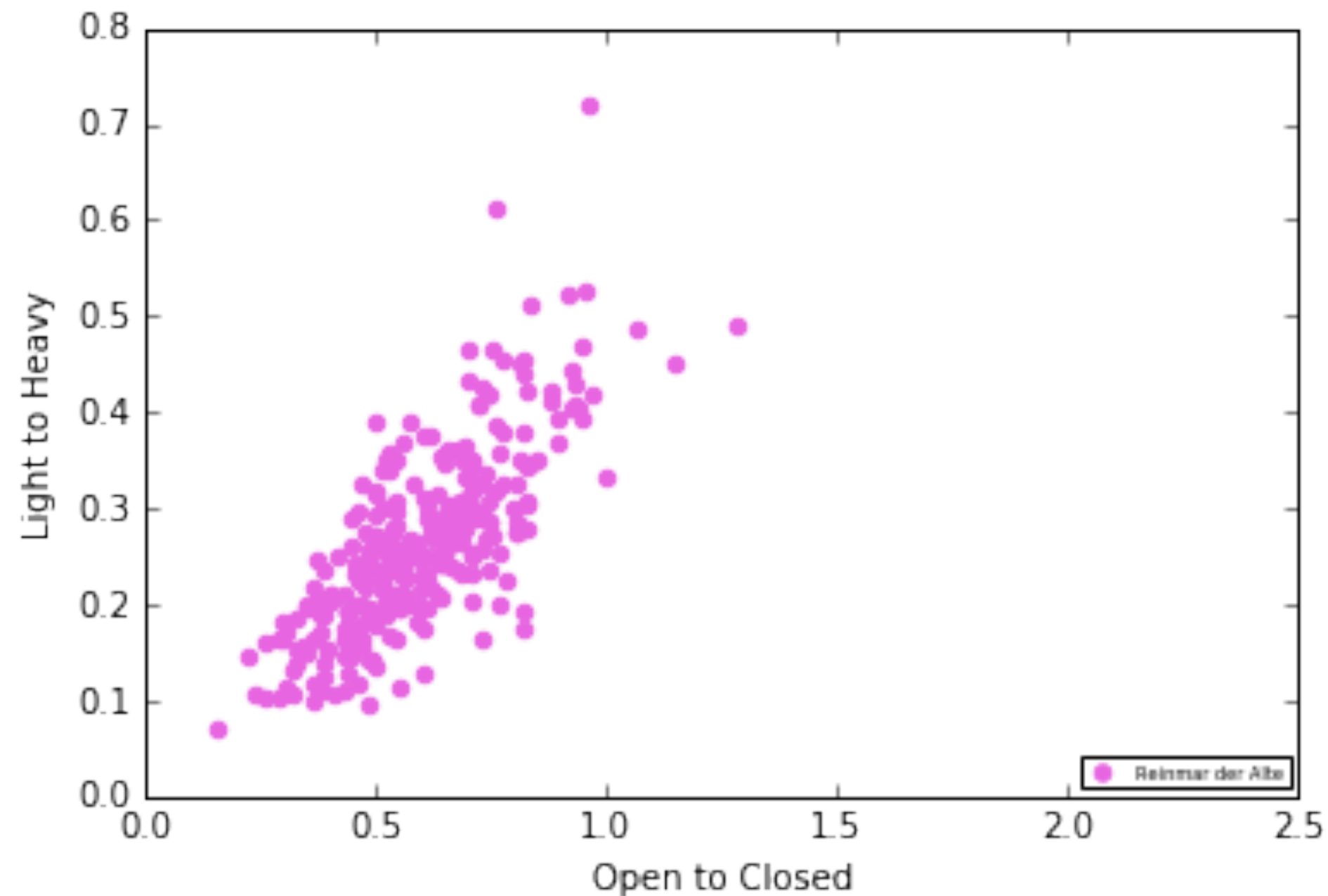
The Sound of Minnesang



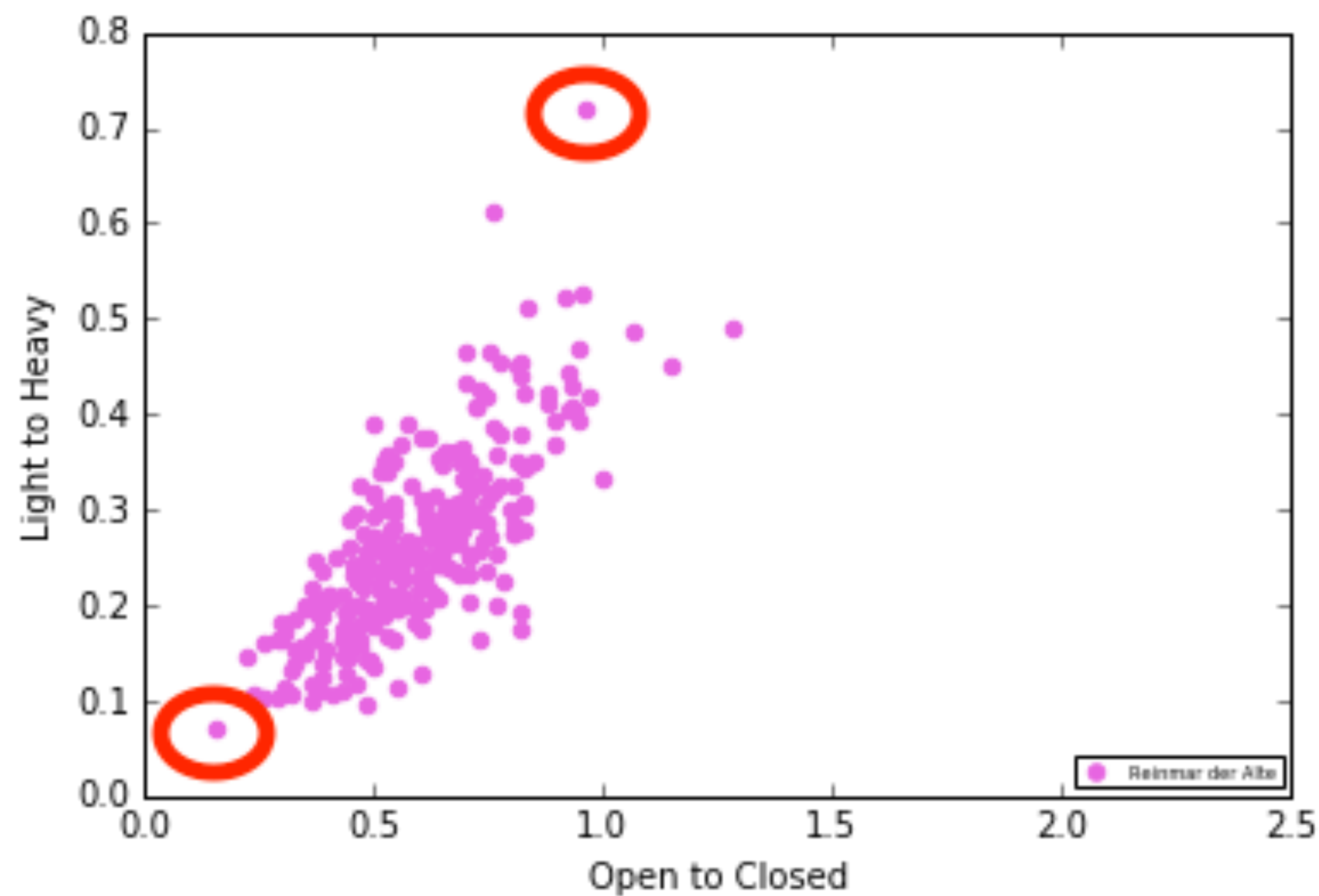
The Sound of Minnesang



The Sound of Minnesang



The Sound of Minnesang



Reinmar der Alte

(192,25-214C)

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

47 word(s)

59 syllable(s)

1.25531914894
syllable(s) per
word

4 light syllable(s)

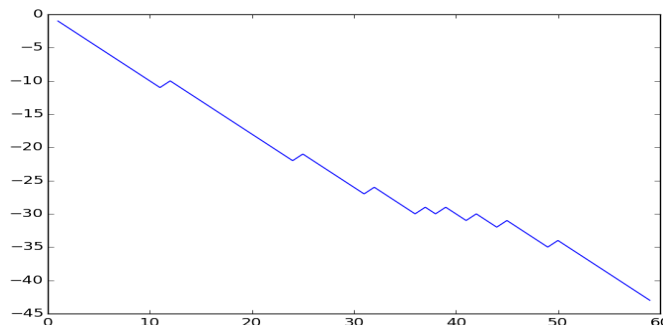
55 heavy
syllable(s)

6.77966101695
percent light

9 open syllable(s)

50 closed
syllable(s)

15.2542372881
percent open



Reinmar der Alte

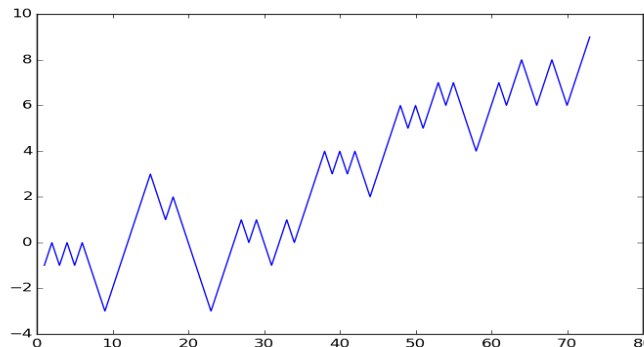
(173,13-96C,53b)

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

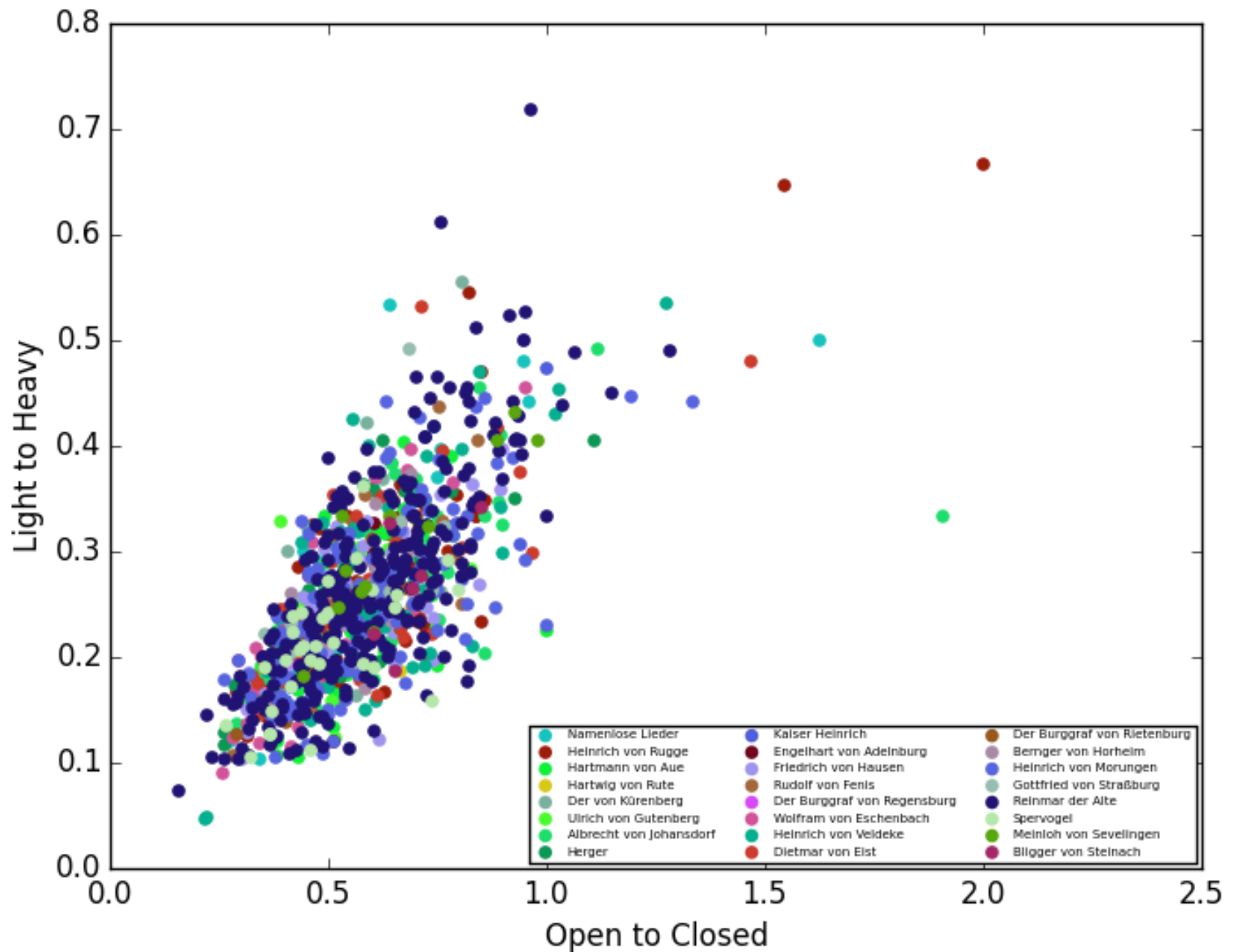
46 word(s)
73 syllable(s)
1.58695652174
syllable(s) per
word

25 light syllable(s)
48 heavy
syllable(s)
34.2465753425
percent light

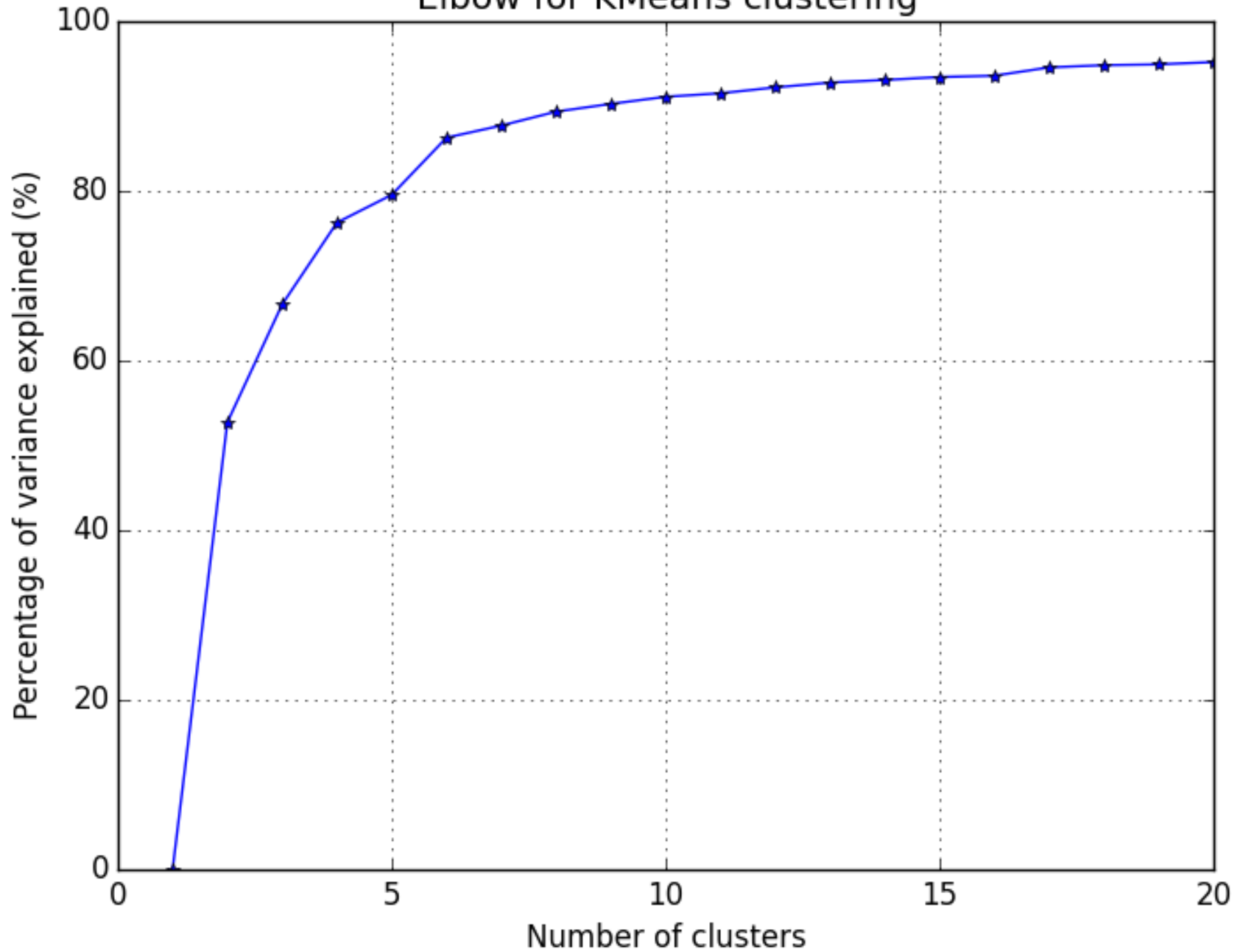
42 open syllable(s)
31 closed
syllable(s)
57.5342465753
percent open



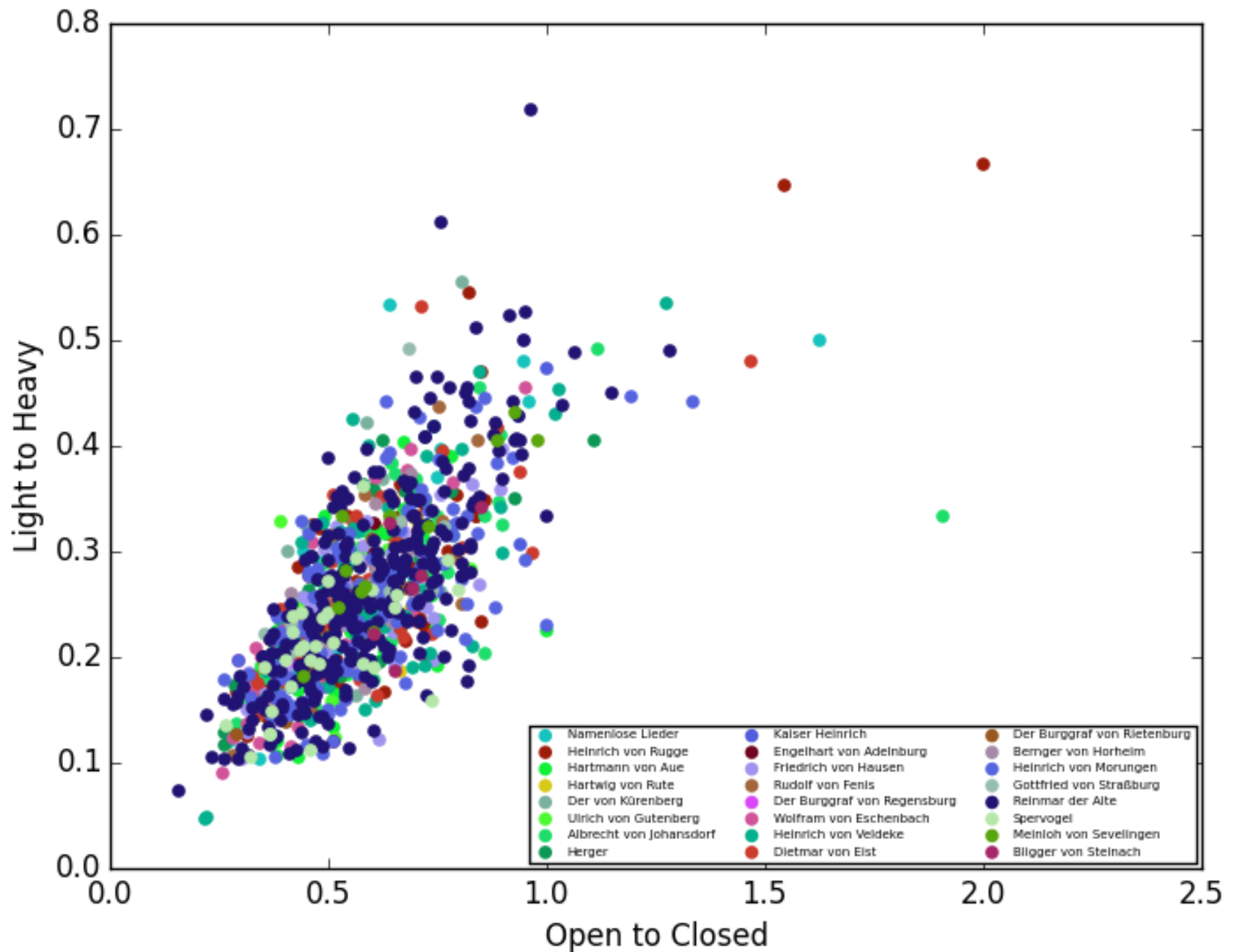
The Sound of Minnesang



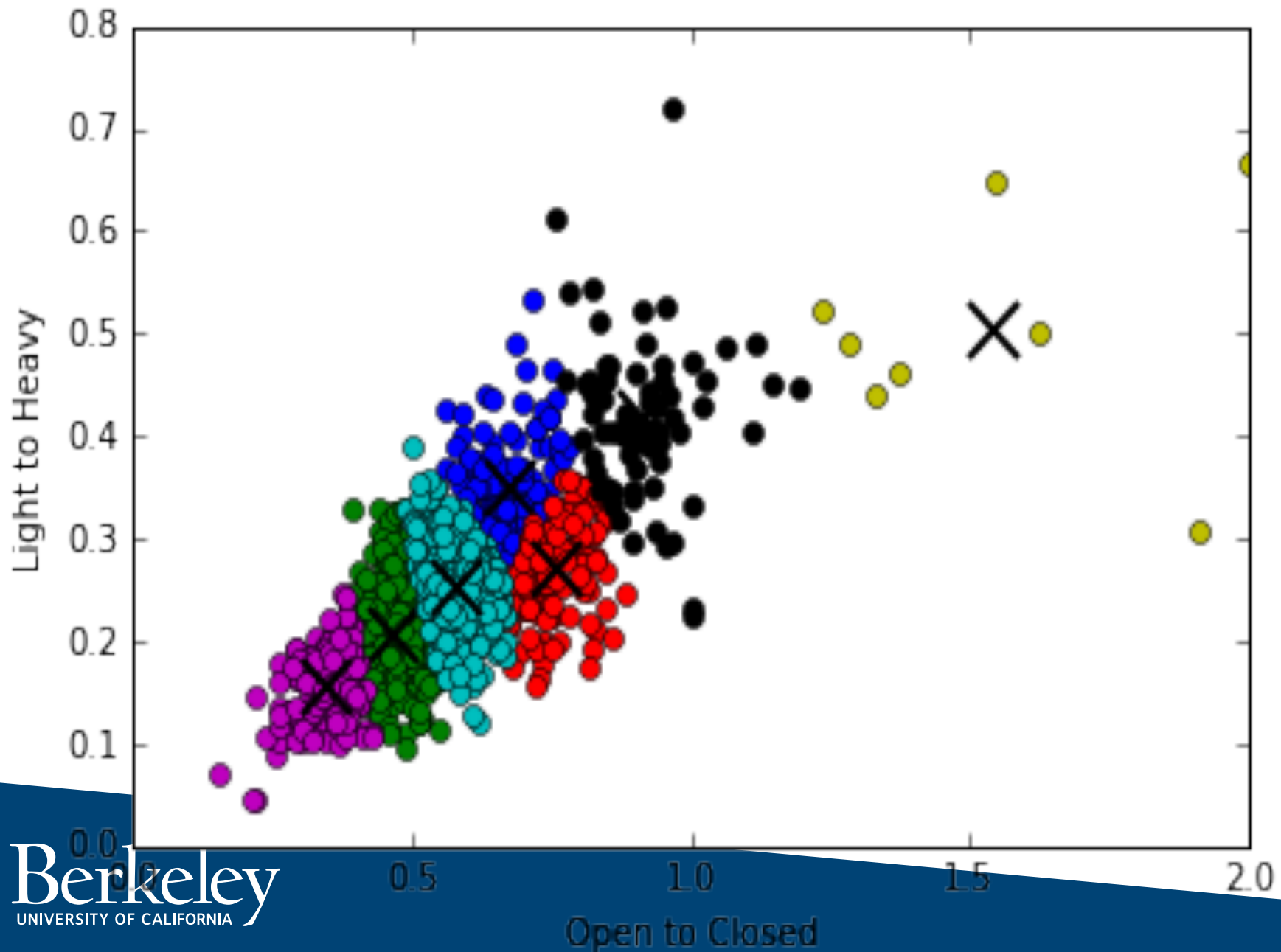
Elbow for KMeans clustering

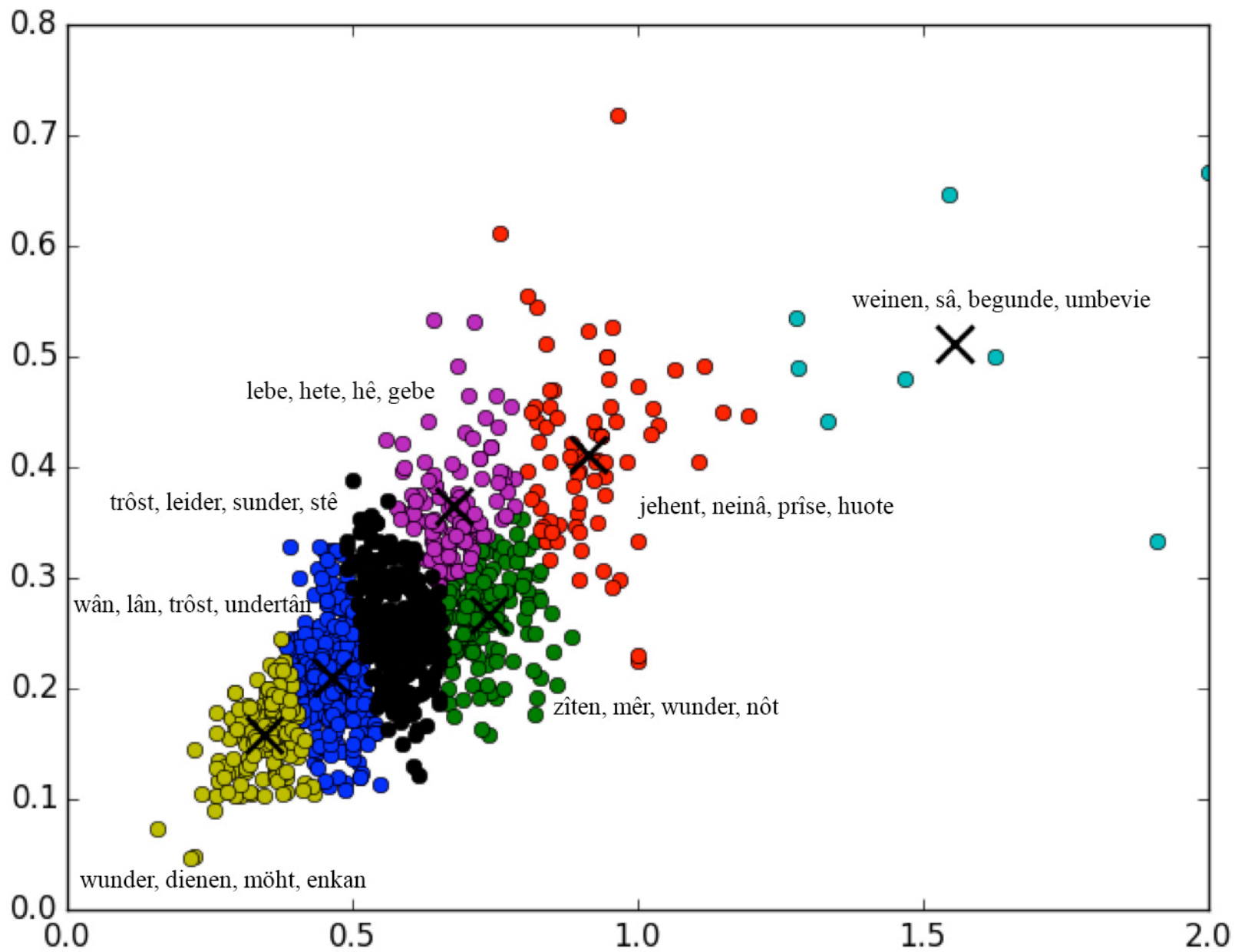


The Sound of Minnesang



DMF Sound Clusters





Heinrich von Morungen 143,22-93C ff

1, Owê, —
2 sol aber mir iemer mê
3 geliuhten dur die naht
4 noch wîzer danne ein snê
5 ir lîp vil wol geslaht?
6 der trouc diu ougen mîn.
7 ich wânde, ez solde sîn
8 des liechten mânen schîn.
9 dô tagte ez.

1 2, Owê, —
2 sol aber er iemer mê
3 den morgen hie betagen,
4 als uns diu naht engê,
5 daz wir niht durfen klagen:
6 owê, nu ist ez tac,
7 als er mit klage pflac,
8 dô er jungest bî mir lac?
9 dô tagte ez.

1 3, Owê, —
2 si kuste âne zal
3 in dem slâfe mich.
4 dô vielen hin zetal
5 ir trehene nider sich.
6 iedoch getrôste ich sie,
7 daz si ir **weinen** lie
8 und mich alumbe vie.
9 dô tagte ez.

1 4, Owê, —
2 daz er sô dicke sich
3 bî mir ersehen hât!
4 als er endachte mich,
5 sô wolt er sunder wât
6 mîn arme schouwen blôz.
7 ez was ein **wunder** grôz,
8 daz in des nie verdrôz.
9 dô tagte ez.

Skandierung

“Hybrid Meter” (akzentbasierte Quantitätssprache)

- Klassische Metrik
- mittelalterliche Varianten
- Auflösung

Quantitative Metrik

Aeneid

āarma vi |rumque ca|nō, Troi |ae quī|prīmus ab|ōrīs
— ^ ^ | — ^ ^ | — — | — — ^ ^ | — —

Qualitative Metrik

Romeo and Juliet 1.1.1:

Two house | holds, both | alike | in dig|nity. |
| × ×' | × ×' | ××' | × ×' | × ×' |

MHD Metrik

- mittelhochdeutsche epische Metrik
- qualitativ und quantitativ
- Abwechslung von Hebungen und Senkungen

MHD Metrik

Der Arme Heinrich:

Ein rîter sô gelêret was

X | X' X | X' X | X' X | X' ^

daz er an den buochen las

| X' X | X' X | X' X | X' ^

swaz er dar an geschriben vant

X | X' X | X' X | X' X | X' ^

Der was Hartmann genant

| X' X | --- | X` X | X' ^

dienstman was er zeOuwe

| X' X | X' X | --- | X` ^

MHD Metrik

- Ein Versfuß kann aus einer Silbe bestehen, die phonologisch lang ist (Doppelmora)
- Eine lange Silbe hat einen langen Vokal oder endet mit einem Konsonant

MHD Metrik

- Ein Versfuß kann aus drei Silben bestehen
- Zwei davon müssen phonologisch kurz sein (Halbmora), die zusammen als eine Silbe wirken
- Eine kurze Silbe muss mit einem kurzen Vokal enden

MHD Metrik

der nam im manege schouwe (6)

X | X' X | ^' ^ X | — | X` ^

- Der zweite Versfuß besteht aus drei Silben
- Der dritte Versfuß besteht aus einer Silbe.

MHD Metrik

- Haupthebung: die erste oder einzige Betonung eines Wortes
- Nebenhebung: eine folgende betonte Silbe im selben Wort
- Die anderen Silben sind unbetont

der nam im manege schouwe (6)

X | X' X | ^' ^ X | — | X` ^

MHD Metrik

- Ein abschließendes unbetontes “e” kann ausgelassen werden, wenn das nächste Wort mit einem Vokal anfängt

darán begundē er suochen (8)

X|X' X | X' X | — | X` ^

Literatur

- Hartman, 1996
- Plamondon, 2006; McAleese, 2007; Greene et al., 2010; Agirrezabal et al., 2013; Navarro, 2015
- Dimpel, 2002

Skandierungsprozess



Metrische Werte

- Die 8 metrische Werte, die gelernt werden müssen:
 1. Mora – Haupthebung
 2. Mora – Nebenhebung
 3. Mora – Senkung
 4. Halbhora – Haupthebung (gespalten)
 5. Halbhora – Nebenhebung (gespalten)
 6. Halbhora – Senkung (gespalten)
 7. Doppelmora – beschwerte Hebung
 8. Auslassung

Daten

	mean	std.	min.	max.
char. per line	21.34	3.39	9	32
syll. per line.	7.62	1.04	5	11
words per line	5.01	1.13	1	8
char. per word	4.26	1.96	1	17
syll. per word	1.52	.71	1	7
char. per syll.	2.80	.81	1	7

Table 1: Summary statistics for annotated dataset

Markierungen

ein/MORA WBY/WBY rî/MORA_HAUPT ter/MORA WBY/WBY sô/
MORA_HAUPT WBY/WBY ge/MORA lê/MORA_HAUPT ret/MORA WBY/
WBY wäs/MORA_HAUPT

daz/MORA_HAUPT WBY/WBY er/MORA WBY/WBY an/MORA_HAUPT
WBY/WBY den/MORA WBY/WBY buo/MORA_HAUPT chen/MORA WBY/
WBY las/MORA_HAUPT

swaz/MORA WBY/WBY er/MORA_HAUPT WBY/WBY dar/MORA WBY/WBY
an/MORA_HAUPT WBY/WBY ge/MORA schri/MORA_HAUPT ben/MORA
WBY/WBY vant/MORA_HAUPT

der/MORA_HAUPT WBY/WBY was/MORA WBY/WBY hart/DOPPEL man/
MORA_NEBEN WBY/WBY ge/MORA nant/MORA_HAUPT

dienst/MORA_HAUPT man/MORA WBY/WBY was/MORA_HAUPT WBY/
WBY er/MORA WBY/WBY zuo/EL ou/DOPPEL we/MORA_NEBEN

Inter-annotator agreement

		Annotator 2							
		×	×	×	—	∪	∩	e	∩
Annotator 1	×	285	4	0	1	3	0	0	0
	×	0	225	1	0	1	0	0	0
	×	1	2	74	0	2	0	0	0
	—	1	2	0	72	0	0	0	0
	∪	1	0	0	0	36	0	0	0
	∩	0	1	0	0	0	17	0	0
	e	0	0	0	0	0	0	9	0
	∩	0	0	0	0	0	0	0	1

Table 2: Inter-annotator agreement confusion matrix

Modelle

- N-gram
- Brill Tagger
- Conditional Random Fields

CRF Model

- Stelle in der Zeile
- Länge der Silbe (Buchstaben)
- Ausschnitte
- Auslassung
- Silbengewicht und –länge
- Wortgrenzen

Ergebnisse

metrical value	F	obs.	held-out	
			F	obs.
mora - unstr.	.922	2403	.938	253
mora - prim.	.918	2025	.947	216
mora - sec.	.856	437	.880	37
double mora	.842	425	.865	34
half mora - unstr.	.574	231	.585	41
half mora - prim.	.771	103	.737	11
elision	.763	65	.500	2
half mora - sec.	0	4	0	0
(macro) average	.894		.904	

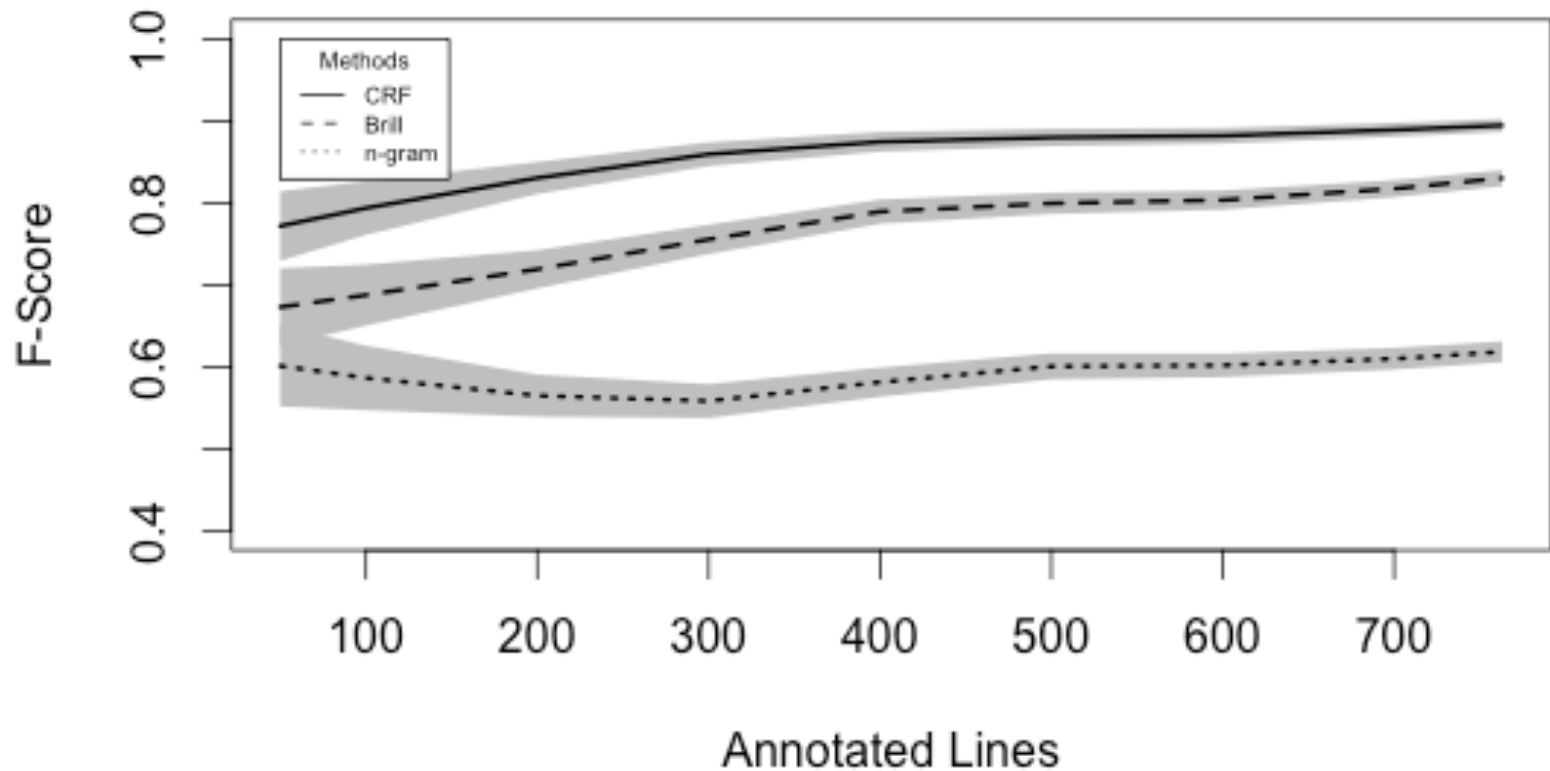
Table 3: CRF model F-score for individual metrical values and (macro) average in development and on held-out data

Probleme

und dâ mite er sich möhte²⁰

CRF model: | x̂ x | x̂ x | x̂ x | x̂ ^
correct: x | x̂ x | x̂ x | — | x̂ ^

Accuracy with Additional Annotated Data



Regeln

CRF	Brill
(1) not \acute{x} if next syll. is end of line	(1) $\acute{x} \rightarrow \times$ if at word boundary and following syll. is \acute{x}
(2) — if end of line is next syll.	(2) $\acute{x} \rightarrow \text{—}$ if followed by \grave{x} and word boundary
(3) e^{\cdot} if last char is “e” and first char. of next syll. is “e”	(3) $\times \rightarrow \grave{x}$ if end of line
(4) not — if syll. is open and light	(4) $\grave{x} \rightarrow \acute{x}$ if monosyllabic
(5) \times if syll in pos. +7 is not end of line	(5) $\times \rightarrow \acute{x}$ if following syll. is “ge”

Table 4: Top five CRF features and Brill rules

502: von a-râ-bîe des gol-des

X / X' X / X` X / ---' / X` Zweisilbig klingend

503: he-ter ma-ne-gen knol-len brâht

/ X' X / X' ∪ ∪ / X' X / X' Einsilbig männlich

504: liu-te vin-ster sô diu naht

/ X' X / X' X / X' X / X' Einsilbig männlich

505: wârñ al-le die von za-za-manc

X / X' X / X' X / X' X / X` Dreisilbig klingend

506: bî den dûht in diu wî-le lanc

/ X' X / X' ∪ ∪ / X' X / X' Einsilbig männlich

zukünftige Arbeit

Wolfram – *Parzival* and *Willehalm*

Hartmann – *Iwein*, *Erec*, *Der arme Heinrich*, *Gregorius*

“wârheit”:

- Wolfram = 8/51 (15.7%)
- Hartmann = 28/33 (84.8%)

“minnen”:

- Wolfram = 8/73 (11.0%)
- Hartmann = 13/16 (81.3%)

Source code

- <https://github.com/henchc/>
- syllabipy.com