

# Variable Behavior of English Predicate Negation

and its implications for acquisition

Phillip Potamites

May 1, 2008

## Abstract

Although verb-negation sequencing has been analyzed as a discrete parameter, this study reports on its variable manifestation in *Moby Dick*.<sup>1</sup> Meanwhile, though contraction is an evidently optional phenomenon, it approaches near determinism in *Huckleberry Finn*. Thus, arguments from Zwicky and Pullum (1983)'s syntactic idiom analysis of auxiliary-negation contraction (ANcon) are used to emphasize non-modularized, *a posteriori*, probabilistic requirements of the acquisition process. While shortcomings of simple-minded string frequency acquisition are acknowledged (cf. Krug (1998)), proceeding from imitation to generalization would seem to require both compositional and decompositional capabilities, conditioned by phonological, lexical, syntactic class, and string frequency effects, in both ontogenic and phylogenetic development.

## 1 Introduction

This introduction summarizes data collected, data excluded, and distributions of its 2 dependent variables (LVP & ANcon, see the relevant sections for definitions) by sentential constructions.

This study focused mainly on 1352 manually analyzed examples of “not” and “n’t” occurring in Herman Melville’s *Moby Dick*. Additional, automatically-extracted comparisons<sup>2</sup> are made to Mark Twain’s *Huckleberry Finn* and (briefly<sup>3</sup>) Robert Louis Stevenson’s *Treasure Island*. Though all three authors travelled widely, both in life and fiction, their styles should be considered for regional differences at approximately the same time, as shown by the table below. The topics of sequencing and contraction addressed here, however, are probably more personally stylistic, than regional.

	s.	p.	year	origin
Moby Dick	9121	37916	1851	New York
Huck Finn	6368	18014	1884	Missourri
T. Island	3793	12038	1883	Scotland

---

<sup>1</sup>All literary, grammatical quotes should be assumed to come from this text, unless otherwise noted.

<sup>2</sup>There are small discrepancies in the manual vs. automated counts, because the manually analyzed selection was originally case-sensitive. In the future, this inconsistency needs to be remedied.

<sup>3</sup>Future research will hopefully extend these comparisons, and incorporate Shakespearean precedents.

The sentence count (s.) is an automatically generated approximation, while the phrase count (p.) indicates any punctuation break. Given the not-uncommon variable use of every sort of punctuation, from periods, to semi-colons, commas, and dashes, as either sentence breaks or not, I've begun to rely on the phrase count, as more relevant and dependable.

The scope of this study is restricted by orthographic conventions. Hopefully, these writing tendencies can be considered somewhat indicative of linguistic faculties, but it should not be presumed that coalescence only manifests where orthographically indicated (also see Mohr (1969) for literary evidence of contraction's under-representation.).

Since this study focuses on negation's interaction with verbs, and particularly auxiliaries, it seems appropriate to exclude "not" where it directly modifies a noun or adverb, as illustrated by the following examples:

- (1) a. Such dreary streets ! blocks of blackness , not houses , on either hand , and here and there a candle , like a candle moving about in a tomb .
- b. He really did this in not only a civil but a really kind and charitable way .

Non-finites also appear to pre-empt the need to fill the pre-negative position, thus also pre-empting both the possibility of "n't" suffixation and any necessary alternation between auxiliaries and lexical verb precedence (LVP).<sup>4</sup>

(3) no aux, neg-initial:

- a. For what seemed ages piled on ages , I lay there , frozen with the most awful fears , not daring to drag away my hand ;
- b. So he makes the best of it ; and when the sailors find him not to be the man that is advertised , they let him pass , and he descends into the cabin .

The particular requirements of finite clauses well might be taken to imply something about correlations of tense and this pre-negation position (though non-finites also show variability, see fn. 4); more relevantly, in looking at where this position is necessarily filled, by verbs or auxiliaries, we necessarily focus on finite contexts.

Moby Dick			/1352			/1012		
sentential	1012	.748	finite	894	.883			
nominal	120	.089	other	118	.117			
adverbial, etc.	220	.163						

Thus, the relevant data is reduced from 1352 occurrences, to 1012, to 894.

The significance of pronominal arguments, tense, phonological environments, and other variables will be presented where relevant below.

To summarize and set aside exclusively syntactic concerns:  $\chi^2$ -tests reflect a significant impact of sentential typology on both contraction and LVP (fig. 1).

<sup>4</sup>However, *Moby Dick* does display LVP in non-finite contexts.

- (2) a. Supper over , the company went back to the bar - room , when , knowing not what else to do with myself , I resolved to spend the rest of the evening as a looker on .
- b. But to my astonishment , he sat down again on the transom very quietly , and seemed to have not the slightest intention of withdrawing .

LVP: $p(\chi^2 > 118.3, f^\circ = 3) \approx 0$				
MD	ind.	qu.	imp.	hyp.
VN	64	8	32	3
NV	580	137	24	46

ANcon: $p(\chi^2 > 50.8, f^\circ = 3) \approx 0$				
MD	ind.	qu.	imp.	hyp.
n't	156	73	22	5
not	488	72	34	44

Figure 1: Sentential Type for ANcon and LVP

Both LVP and ANcon are highly unlikely in hypothetical contexts<sup>5</sup>, pretty likely in imperative contexts<sup>6</sup>, and questions almost lean towards ANcon<sup>7</sup> while pretty strongly discouraging LVP. Indicative contexts for both phenomena, in this particular novel, might be characterized as healthy minorities; from close to 10% (VN | neg.ind.) to over 20% (n't | neg.ind.), they are not dominant grammatical requirements, but they are hardly marginalized nor necessarily destined for demise.<sup>8</sup>

## 2 LVP was variable in some American dialects of the late 1800's. (Melville)

Lexical Verb Precedence (LVP) is supposed to be a more theory-independent name for what is commonly known as “verb-raising”. The idea behind ‘verb-raising’ is that the verb has originated in a position subordinate to the negation and then ‘raised’ to a superior position, thus ultimately preceding the negation, as in the following examples:

- (6) a. He sees no black sky and raging sea , feels not the reeling timbers , and little hears he or heeds he the far rush of the mighty whale , which even now with open mouth is cleaving the seas after him .
- b. Talk not that lingo to me .
- c. Perchance ye need it not .<sup>9</sup>
- d. This lovely light , it lights not me ;

In contemporary English, it is frequently assumed that only the small set of items known “auxiliaries” can occupy this pre-verbal position.

<sup>5</sup>Problematically, this category includes both of the following kinds of clauses:

- (4) a. Had not the stranger stood between me and the door , I would have bolted out of it quicker than ever I bolted a dinner .
- b. If he had not been a small degree civilized , he very probably would not have troubled himself with boots at all ;

<sup>6</sup>5/32 of the LVP imperatives counted ‘let’ as a lexical verb. As Zubizaretta notes (p.c.), the following sentences are hardly (currently?) equivalent:

- (5) a. Let him not go. ≠
- b. Don't let him go.

This issue raises a question of ‘porousness’ in negation interpretation. See fn. 18 on differences between *can* and *must*.

<sup>7</sup>Zwicky and Pullum (1983) cite negative contractions in questions as proof of ‘syntactic unity’.

<sup>8</sup>Though a contemporary comparison would probably suggest that, currently, LVP is predominately lower and ANcon higher than their representations in *Moby Dick*.

<sup>9</sup>Is the pronoun acting like a clitic? Hopefully future research will capture pronominal, positional, and categorical factors of verb-phrase material.

Different word orders in basically equivalent, French and English sentences, like those below, have led researchers to describe variation across languages as binary parameters deriving properties like LVP (for instance, see Pollack (1989), Roberts (1993), Chomsky (1986), or Rizzi (1990)).

- (7) a. Je lis souvent des livres.  
b. I often read books.
  
- (8) a. Je ne lis pas des livres.  
b. I do not read books.

The French verb ('lis') is considered "raised" because it proceeds the negative 'pas' and/or the adverb 'souvent'.

However, there are a variety of complications to bear in mind: because of (at least) the distinct behavior of french infinitives, and *etre* and *avoir* (as infinitives), and modals more generally, Pollack (1989) implicates tense,  $\theta$ -assignment, and modal base generation in accounting for the data. In accounting for English sequencing, Roberts (1993) primarily implicates parameter settings including  $\text{Agr}^{-1}$  and  $\text{T}^{-1}$ , without entirely explicating exactly what he assumes about  $\text{X}^{-1}$  (though it has something to do with defective inflectional paradigms; see his p. 244).

Others note that negation and adverbs don't necessarily pattern identically. Second language acquisition researchers like Lydia White and Lynn Eubank both assume that this parameter is off in target English, while, respectively, finding or assuming adverb-negation distinctions. Eubank argues that negation LVP is less likely than adverbial LVP, which runs contrary to my own intuitions.

- (9) "If the negative morpheme heads NEGP (rather than being its specifier), then verb raising past negation cannot take place without violating the ECP (see, e.g. Zanuttini 1991). In other words, it appears that employing a contrasting pair like [ex. 9a-b] will not, of necessity, implicate the inflection-raising relation." -Eubank and Grace (1998) (p. 73)

I've instead purposely studied negation LVP, both because of ease, and because it is quite common (in any of the sort of mandatory high school literature from earlier eras), but future research is needed to test my suspicions that negation LVP is more common than adverbial LVP.

In fact, both Eubank and Grace (1998) and White (2003) find that their second language learners develop various tolerances to the VAdv word order, regardless of its (supposed) absence in both L1s, of Chinese and Turkish, respectively, and the L2, English.

Eubank and Grace (1998)'s advanced learners (inflectors) were able to match these word orders quite quickly, without the grammatical facilitation that gave both natives and beginners (non-inflectors) lower RTs with the AdvV order.

z, ms	VAdv	AdvV
non-inflectors	.1067, 3038	-.1040, 2841
inflectors	-.0360, 2594	.0353, 2618
native	.0773, 1546	-.0762, 1491

White first observes “some cases” where lexical verbs precede adverbs in spontaneous production.

“However, there are some cases where the lexical verb is placed to the left of adverbs and prepositional phrases, as can be seen in [10=her ex. 15].

(10) That’s what I notice because my daughter speaks very well English.” -White (2003) (p. 135)

Those production errors were then further substantiated by (off-line) judgement tasks which are reported in table 1.

Table 1: “correct rejections of verb raising”

White (2003)	(table 7, p. 135)
questions	7/7
negatives	4/4
adverbs	2/4

The data from *Moby Dick* is not intended to suggest that these learners acquired this construction from literary exposure (especially since I examined negation, and haven’t yet looked for adverb evidence), but it does qualify any claim that English necessarily lacks verb raising, especially if negative contexts are supposed to be particularly bad (like potentially violating the ECP, as in ex. 9).

Regardless of the numerous, necessary qualifications of this phenomenon, Yang (1999)’s account basically advocates a deterministic property in a probabilistic set of grammars:<sup>10</sup>

“For example, sentences that contain finite verbs preceding adverb or negation (“Jean voit souvent/pas Marie”) are unambiguous indication for the [+] value of the verb raising parameter. A grammar with the [-] value for this parameter is incompatible with such sentences and if probabilistically selected for the learner for grammatical analysis, will be punished as a result.” -Yang (1999), p. 43.

A ‘competing grammar’ analysis of optionality may be feasible as a generative decision procedure. However, accurately emulating grammatical intuitions (if so) doesn’t necessarily prove such a procedure’s role in human behavior. Yang more persuasively correlates statistical distributions of triggers that he (perhaps too subjectively) identifies in the data with acquisition age. Nevertheless, there is no clear empirical justification suggesting that the graded correlations of phonology,

<sup>10</sup>Though he recognizes the complexity problem of this perspective:

“According to some estimates ([...]), 30-40 binary parameters are required to give a reasonable coverage of the UG space. And, if the grammars are stored as individual wholes, the learner would have to manipulate  $2^{30}$ - $2^{40}$  grammar weights: now *that* seems implausible.” - Yang (2003)

He also acknowledges the problems with the deterministic view of parameters (Yang (2000), p. 233):

- “However, this discrete view of language learning is at odds with the facts of language development.”
- “In general, child language development is gradual, thus providing no support for the on/off view of parameter setting”.

syntax, and lexicon reported here are evidence of totally distinct systems at work. All the same, these graded factors are probably effective in both ontogenic and phylogenetic development.

The development of non-target behaviors, as with other non-input outputs, is an exciting research topic, which is set aside here, in attempting to first ascertain characteristics of available, empirical evidence (as both potential inputs and attested outputs).

Finally, before presenting the data, I need to confess that, though it is easy to distinguish nominal negation (as done in the intro) in the subject position, the distinction is more debatable in the object position:

- (11) a. Not a man spoke.  $\neq$   
b. A man did not speak.
- (12) a. He heard not a word.  $\approx?$   
b. He did not hear a word.

Ex. 12b is most informative with wide negation (there is no word that he heard), but also can allow the wide indefinite reading (there is a word that he didn't hear). More research is needed to determine if this ambiguity also existed in LVP structures (looking like 12a). Presumably nominal negation (also looking like 12a), with "not a" behaving like a quantifier, would only permit the wide negation reading. Thus, either analysis should provide the most salient reading (of 12a), but only predicate negation should permit the ambiguity.

In debatable cases, I've erred in favor of LPV, and counted them as VN structures. Note that quantifier phrases (QPs) like "not a" or "not one" are far more natural (for my current dialect) than ones like "not the", "not some", or especially "not what".<sup>11</sup> Between many such unlikely QPs and a notable frequency of intransitives<sup>12</sup>, these questionable inclusions are unlikely to be an excessive proportion. More precise coding strategies are needed in the future to distinguish these groups and their semantic overlaps, or the lack thereof.

The tables in fig. 2 document counts and proportions of auxiliary use and LVP. In the left table we see that the likelihood of *do* and other auxiliaries increases in finite contexts while auxiliary absence ( $\emptyset$ ) decreases, as suggested in the introduction.

Auxiliary absence ( $\emptyset$ ) includes most VN structures and thus is a higher percentage in the pred contexts (.149 (left) vs. .119 (top-right, left col.)), where also including non-finite neg-initial structures, but lower than the VN percentage in finite contexts (.087 vs. .12): VN also includes at least the Semi-Auxes, such as *know*, *dare*, and *let*, and even some Aux, such as *have* when used as a possessive. Note that the LVP rate, in fact, remains practically constant.

Restricting our attention to only non-modal contexts where VN is presumed to be specifically interchangeable with *do*-support, we find *do* appearing in about 72% of contexts while lexical verbs

---

<sup>11</sup>Try them in subject position:

- (13) ?Not the/some/what man spoke.

However, though I find the following sentence a bit odd, it suggests an ability to accept nominal 'not' with definites (i.e. the proper name):

- (14) Sway me up , men ; ye have been deceived ; not Moby Dick casts one odd jet that way , and then disappears .

<sup>12</sup>Intransitivity appears to be a crucial factor for which suitable coding schemes need to be determined in the future. Ellegård (1953) also notes the significance of this factor in the development of *do*-support.

	pred	/1012	finite	/ 894	LVP: pred	/1012	finite	/894	
<i>do</i>	284	.248	276	.309	NV	892	.881	787	.88
Aux	542	.536	511	.571	VN	120	.119	107	.12
Semi-Aux	35	.035	29	.032	non-modal				
∅	151	.149	78	.087	p(do   ind,fin,neg)	=	.721		
					p(VN   ind,fin,neg)	=	.279		

Figure 2: Pre-neg fillers and LVP

precede negation (VN) in almost 28% of these contexts (fig. 2, right-bottom).<sup>13</sup> These non-modal contexts make up about 43% of *Moby Dick*'s finite occurrences ((107+276)/894). In sum, LVP was a not uncommon alternation in non-modal negation.

Surprisingly, a  $\chi^2$ -test indicates that Melville's use of LVP is not conditioned by any dialect differences between the narrator's voice and those that he chooses to quote directly (this will not be the case for contraction). The differences are estimated to have a a better than 60% likelihood of being due to chance.

Table 2:  $p(\chi^2 > .248, f^\circ = 1) = .618$

MD	narrator	quotation
VN	56	51
NV	432	355

On the other hand, present tense, in fact, promotes LVP (i.e. VN) (16% of presents vs. 8% of pasts).<sup>14</sup> Why this should be so remains uncertain.

Table 3:  $p(\chi^2 > 8.996, f^\circ = 1) = .0027$

MD	past	present
VN	25	77
NV	278	418

Finally, and also surprisingly, pronominal subjects also encourage LVP (VN, table 4) (13% of pronoun subject sentences vs. 8% of non-pronominal subjects). If string frequency effects contribute to the likelihood of certain constructions (as considered for contraction), we would probably expect *pronoun-auxiliary* strings to be preferable to *pronoun-lexicalVerb*. It could be reasonable to speculate about a factor of *prosodic dissimilation*, but such an effect also requires further research.<sup>15</sup>

<sup>13</sup>These numbers are calculated via the (Bayesian) definition of conditional probability. In our case, 'x' is one possible manifestation, let's call it 'a', and 'y' either of these **complementary** manifestations, call that 'a or b':

$$p(x|y) = \frac{c(x \text{ and } y)}{c(y)} = \frac{c(a \text{ and } (a \text{ or } b))}{c(a \text{ or } b)} = \frac{c(a)}{c(a) + c(b)} = p(a|(a \text{ or } b)), \quad \text{e.g. } \frac{107}{107 + 276} = .279 \quad (1)$$

<sup>14</sup>Note that these counts exclude futures, since they require auxiliaries and categorically exclude VN.

<sup>15</sup>Labov (1969) also notes the pronominal subject effect facilitating subject-auxiliary contraction (SAcon) saying, "The single most important constraint on deletion in NNE, and upon contraction in SE and NNE, is one which we did not expect: whether or not the subject is a pronoun or some other noun phrase" (p. 730). Could SAcon and LVP both be satisfying similar needs for prosodic contrasts? It's probably too optimistic to expect them to be exactly the same.

Table 4:  $p(\chi^2 > 4.35, f^\circ = 1) = .037$

MD	nom subj	pron subj
VN	19	88
NV	214	573

As acknowledged, the optionality of LVP in Melville’s dialect(s) does not directly disprove anything about the generative enterprise. However, if the optionality is obtained by the application of multiple grammars, then the data above suggests that grammar selection may be affected by tense and argument pronominality. Alternatively, tense and argument manifestation are internal to the processes affecting syntactic decisions.

### 3 ANcon was nearly invariable in American dialects of the late 1800’s. (Twain)

Auxiliary-Negation contraction (ANcon) is predominately restricted to auxiliaries, though note that Zwicky and Pullum (1983) include “daren’t” in their considerations. Melville also uses such a form in one of his quotes:

- (15) Who ‘s afraid of him , except the old governor who daresn ‘t catch him and put him in double - darbies , as he deserves ,

ANcon in *Moby Dick*, is also affected by tense, as well as phonological factors, but in *Huck Finn*, the effects are not statistically provable because of the over-riding likelihood of ANcon under almost any possible circumstance.

The tables in fig. 3 show that *Huck Finn* has far and away the highest use of negation by phrase (p.), though it barely uses ‘not’ 1% of the time. *Moby Dick*, contra its American cousin, uses contraction a good deal less than the UK data point represented by *Treasure Island*.

Huck Finn	/p.	/neg	Moby Dick	/p.	/neg	T. Island	/p.	/neg			
not	165	.009	.074	not	1142	.030	.806	not	344	.029	.594
n’t	2075	.115	.926	n’t	274	.007	.194	n’t	235	.020	.406

Figure 3: ANcon by count, phrase and combined neg

These dialect distinctions are also evidenced by the significance of quotation on contraction in Melville’s writing style.

Table 5:  $p(\chi^2 > 286, f^\circ = 1) \approx 0$

MD	narrator	quote
n’t	26	230
not	462	176

On the other hand, the excessively low counts of “not” in *Huck Finn* invalidate the outcome of  $\chi^2$ -tests; if done, as in table 6, for preceding phonological environment, they indicate a better than 1/3 chance of the differences being due to chance.

Table 6:  $p(\chi^2 > .78, f^\circ = 1) = .38$

HF	V	C
n't	341	382
not	1	3

*Moby Dick*, on the other hand, demonstrates a strong effect of preceding phonological context (table 7) (“n’t” occurs in 53% of contexts with preceding Vowels vs. 11% following Consonants). Note that these contexts have been determined strictly on the basis of surface forms, without any presumptions about underlying representations (so “won’t” is counted as a V contraction, though “will” could be considered to end in C.)

Table 7:  $p(\chi^2 > 174, f^\circ = 1) \approx 0$

MD	V	C
n't	163	93
not	117	516

As with LVP, present tense also promotes the likelihood of contraction (table 8) (30% vs. 13%).

Table 8:  $p(\chi^2 > 42, f^\circ = 1) = 1.1 * 10^{-10}$

MD	present	past
n't	176	44
not	319	259

Although the above results are maintained, even when “do” is excluded<sup>16</sup>, “do” is a major proportion of both the V and present “n’t” counts listed above, as shown in table 9 (though these counts, from fig. 4, p. 11, are a little larger because of case-insensitivity).

Table 9:  $p(\chi^2 > 37, f^\circ = 1) = 1.2 * 10^{-9}$

MD	do	did
n't	119	33
not	39	58

However, even where preceding phonological contexts are identical, there are still significant differences between specific lexical items, as shown by table 10.

Classes of lexical items also behave significantly, across tenses and phonology (table 11) (53% of *do* vs. 3% of *be*).

Given the surprising rarity of *be-n't*, one might wonder whether *be* is so discouraged from ANcon because it prefers to undergo subject-auxiliary contraction (SAcon), which *do* generally

<sup>16</sup>Do excluded:

phono: $p(\chi^2 > 66.41, f^\circ = 1) = 3.7 * 10^{-16}$			tense: $p(\chi^2 > 11.56, f^\circ = 1) = 6.7 * 10^{-4}$		
MD	V	C	MD	present	past
n't	52	59	n't	63	13
not	66	436	not	249	198

Table 10:  $p(\chi^2 > 20.59, f^\circ = 2) = 3.4 * 10^{-5}$

MD	did	would	had
n't	33	8	0
not	58	40	36

Table 11:  $p(\chi^2 > 128.3, f^\circ = 1) \approx 0$

MD	do	be
n't	145	6
not	131	189

does not participate in. However, SAcon only accounts for 6% (12/189) of the occurrences of *be* that preferred to appear with “not”. For unknown reasons, the 2 contractions are generally complementary (for a given auxiliary), so that there are no occurrences of “he’sn’t”, though there have been attested items like “wouldn’t’ve” and “I’d’ve” (though not in these books).

On the other hand, *Huck Finn* generally favors ANcon over SAcon (table 12, with ‘-’ standing for neither, e.g. ‘I am not’), so categorically that the  $\chi^2$ -test becomes irrelevant again. Note that this preference is not due to any shortage of SAcon (see fig. 4, p. 11, lower-right).

Table 12:  $p(\chi^2 > 0.0, f^\circ = 10/0) = 1$

HF	SA	AN	--
I-am/ai-not	0	33	0
I-would-not	0	43	0
I-have-not	0	2	0
I-will-not	0	15	0
it-is/ai-not	0	53	0
he-is/ai-not	0	18	0

However, in the case of *Moby Dick*, one might speculate about string frequency effects promoting “it’s” but not “he’s” over ANcon (table 13, also fig. 4, p. 11). However, issues with this kind of analysis are considered in the next section.

Table 13:  $p(\chi^2 > 15.9, f^\circ = 2) = 3.5 * 10^{-4}$

MD	SA	AN	--
it-is/ai-not	5	1	18
he-is/ai-not	0	5	2

Returning to *Huck Finn*, the table at the top-right of fig. 4, p. 11, also indicates that *do-neg* nearly invariably manifests as contraction. Note that I have incorrectly rounded down from 100%. The probabilities derive from various combinations of some of the exact counts sampled below them. Thus,  $p(\text{didn't} \mid \text{did,neg})$ , in fact,  $= 347 / (347 + 1)$ , which is actually much closer to 100%. The rest of the counts in this by no-means-exhaustive selection are presented for any curious reader who might wish to calculate conditional probabilities of their own.

Moby Dick	
p(don't   do,neg)	.75
p(didn't   did,neg)	.36
p(doesn't   does,neg)	.12
p(isn't   is,neg)	.01
p(aren't   are,neg)	.00
p(wasn't   was,neg)	.08
p(weren't   were,neg)	.00
p(it's   it,((i)s∨(ha)s))	.24
p(he's   he,((i)s∨(ha)s))	.50
p(I'd   I,((ha)d∨(woul)d))	.13

Huck Finn	
p(don't   do,neg)	.99
p(didn't   did,neg)	.99
p(doesn't   does,neg)	1.00
p(isn't   is,neg)	1.00
p(aren't   are,neg)	na
p(wasn't   was,neg)	.95
p(wasn't   (wasn't∨warn't))	.07
p(weren't   were,neg)	na
p(it's   it,((i)s∨(ha)s))	.88
p(he's   he,((i)s∨(ha)s))	.84
p(I'd   I,((ha)d∨(woul)d))	.41

Moby Dick 9121 s.			
don't	119	do not	39
didn't	33	did not	58
won't	32	will not	28
ain't	21	was not	58
wasn't	5	has not	8
hasn't	4	have not	9
haven't	3	does not	22
doesn't	3	is not	67
isn't	1	are not	16
aren't	0	were not	12
weren't	0	am not	6

Huck Finn 6367 s.			
didn't	347	did not	1
don't	341	do not	1
ain't	298	will not	0
won't	67	was not	1
wasn't	21	have not	1
warn't	293	has not	0
haven't	7	does not	0
hasn't	6	is not	0
doesn't	1	am not	0
isn't	1		

		it was	290
		it is	279
		it has	32
it's	100	he is	54
he's	97	he has	42
I've	29	I have	110
I'm	27	I am	73
I'd	12	I had	50
you're	6	I would	32
you'd	3	you are	30
he'd	3	you would	43
they're	2	you had	6
she's	2	he had	114
we've	2	he would	62
they've	1	they are	62
you've	1	she is	5
		we have	20
		they have	34
		you have	19

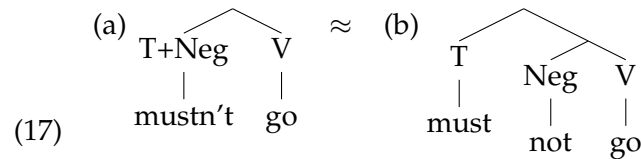
		it was	369
		it is	28
		it has	1
it's	217	I had	88
I'd	98	I would	52
I'm	80	I am	14
he'd	79	he had	70
he's	73	he would	62
I've	49	he is	9
you're	46	he has	5
they're	39	I have	5
you'd	31	you are	22
you've	29	they are	7
they've	25	you would	11
we've	17	you had	11
she's	10	you have	7
		they have	2
		we have	0

Figure 4: con probs and counts

The likelihood of *do-neg*-contraction, in *Moby Dick* (top-left, fig. 4, p. 11), on the other hand, shows a very interesting breadth of variation across its different inflections (from .75 to .36 to .12). Such differences are exactly what manifest statistical significance for both tense and preceding phonological context (because of the consonants in the inflections), but notice that the present tense variable (“don’t” (.75) and “doesn’t” (.12)) patterns on opposite sides of the past (“didn’t” (.36)). These likelihoods do roughly correlate with overall frequencies, but see the next section for qualifications.

Zwicky and Pullum (1983) argue that ANcon is affixation, and thus “part of the syntax proper, not ... post-syntactic cliticization” (p. 511). Paralleling presumed affixes, such as plural -s, past -ed, and superlative -est, and using SAcon as typical cliticization, they point out 1.) n’t is selective in binding with auxiliaries, as each of the affix exemplars respectively binds with nouns, verbs, and adjectives, but unlike “the girl you like’s here.” 2.) There are gaps like the uncertain past participles of ‘dive’ or “stride” and items like “amn’t” or “mayn’t” (though Krug (1998) lists “mayn’t” and *Huck Finn* contains 5 examples), but SAcon is relatively indifferent to it’s host, as the previous example shows. 3.) Morpho-phonological idiosyncrasies exist in the affixes, like irregular verbs and “ain’t” while SAcon manifests as a constant phonological value. 4.) Affixed forms may participate in semantic idiosyncrasies such as “last words”, or “you can’t go home” lacking one of the meanings of “you can not go home”. 5.) Syntactic operations typically move nouns with their plural marking and aux-n’t units can front easily, as in “isn’t it nice?” (cf. “is not it nice?”), while presumably no syntax would treat “like’s” in “the girl you like’s” above as a constituent. 6.) Affixes precede clitics in attachment, and thus, we have “wouldn’t’ve”, but not “would’ve’n’t”. Thus, 3 of their reasons could sound more lexical (a.k.a. arbitrary exceptions) than syntactic.<sup>17</sup>

The goal of the learner, within a theoretical approximation, would seem to be learning some sort of equivalence like ex. 16.<sup>18</sup>



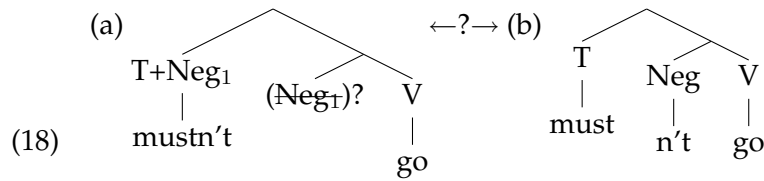
The main point of Zwicky and Pullum (1983) would seem to be that something like ex. 18a should be considered a superior analysis to ex. 18b, where separate terminals are combined at some post-syntactic stage of phonologization.

<sup>17</sup>It may be necessary to reconsider the status of SAcon as well. Zwicky (1970) argues that not only would any sort of SAcon rule require licensing by a [+auxiliary] feature (p. 329), but that it applies to “eight forms only” (p. 328): “is”, “has”, “would”, “had”, “have”, “am”, “are”, and “will”. He notes that it such a rule should not apply to participles, AN forms, and “was” and “were”. Labov more specifically claims “was and were begin with a consonant which is not generally deleted” (p. 721), and later argues “there is probably no general process which removes the w in NNE or SE: a special lexical alternation for will may be required” (p. 755).

<sup>18</sup>“Must” does not seem to allow the same ambiguity as “can”, regardless of imposing similar prosodic patterns.

- (16) a. You can NOT go. ← It’s possible not to go.  
 b. You can not go. ← It’s not possible to go.  
 c. You must NOT go. ← It’s necessary not to go.  
 d. You must not go. ← It’s not necessary to go.

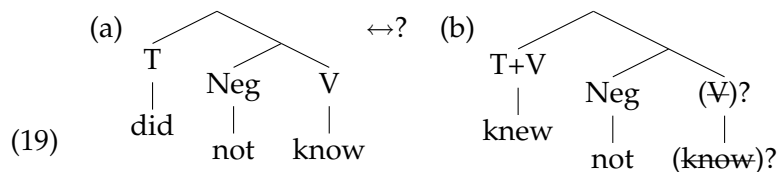
In the case of “can’t”, presumably only the 16b reading is available. Thus, only one interpretation of a structure like ex. 17b should be equivalent to the interpretation of a structure like 17a.



An alternative, however, could be that phonologically simple contractions like “haven’t” undergo simple phonological processes deriving from a form like (b), while suppletive forms like “ain’t” or “won’t” would require a form like (a). Note that /t/ deletion, as in “mustn’t” is disputably “simple”, at least since /d/, in forms like “hadn’t”, doesn’t delete. Though this kind of heterogeneous suggestion may be offensive to a goal of general, uniform rule application, it would also be regrettable to ignore the phonological facilitation that operates in contraction.<sup>19</sup> In either case, at some point, idiosyncratic, lexically-specific exceptions creep in. Thus, I would argue that many rules are *a posteriori* artefacts generalized from E-language exposure, and not absolute, inviolable constraints of human faculties.

The empirical point is that the learner can learn associations like ex. 17, while conclusions about the structures in ex. 18 are primarily theory-internal, notational devices.

We can and should raise the same questions in regard to our earlier, alternative negation manifestation relation, LVP. Melville’s dialect displays a certain level of equivalence between the forms, while others distinguish languages via rule applications that may or may not (necessarily) derive one from the other.



For translation and interpretation, transformational rules associating such alternations are essential. For the learner, however, preferences for one or the other form can be non-deterministically conditioned by lexical item, tense, argument manifestation, perhaps prosodic context, etc.

Thus, while a decision procedure may establish a necessary procedural relation between structures like ex. 19a and 19b or ex. 17a and 17b, a learner must go through an empirical process whereby they do or don’t establish equivalence relations between various strings. Patterns facilitate drawing relations without necessarily deriving from inviolable requirements.

#### 4 String frequency effects require class qualifications.

Krug (1998) argues that “the more frequent a given string of two words is, the higher is its likelihood of contraction” (p. 294). He initially avoids the ambiguity problems of contractions like “it’s” (“it is”/“it has”), by focusing on the contraction “-’ve”, and showing how the probability of the contraction goes up with the sum of its alternations.

By way of comparison, the tables below show how like likelihood of “-’ve” in his data compares with its manifestations in *Moby Dick* and *Huck Finn* (these probabilities can be calculated

<sup>19</sup>The way phonological facilitation competes with semantic needs may be the best account for SAcon’s application to “will”, but not “was” or “were”.

from the counts in fig. 4, p. 11). As with their ANcon behavior, they fall on opposite sides of these UK probabilities. LLC refers to the London Lund Corpus, and BEC refers to the Bank of English Corpus.

Krug (1998): LLC	.5mil w.	Krug (1998): BEC	450?mil w.
p(I've   I,(ha)ve)	.78	p(I've   I,(ha)ve)	.83
p(you've   you,(ha)ve)	.65	p(you've   you,(ha)ve)	.65
p(we've   we,(ha)ve)	.61	p(we've   we,(ha)ve)	.70
p(they've   they,(ha)ve)	.65	p(they've   they,(ha)ve)	.68
Moby Dick	.26mil w.	Huck Finn	.14mil w.
p(I've   I,(ha)ve)	.21	p(I've   I,(ha)ve)	.91
p(you've   you,(ha)ve)	.05	p(you've   you,(ha)ve)	.81
p(we've   we,(ha)ve)	.09	p(we've   we,(ha)ve)	1.00
p(they've   they,(ha)ve)	.03	p(they've   they,(ha)ve)	.93

The graphs on p. 19 present the central data from Krug's 2 corpora 3 different ways. The top 2 tables show, for each corpora, the percentage of "-ve" forms (veper) (like those above) as a function of that contraction's count (ve). The middle tables show just the contraction count (ve) as a function of the sum of its alternatives (tot). The bottom graphs show the contraction count (ve) as a function of the non-contracted alternative count (have). With minor exceptions, the dependent variable (contraction) goes up with the independent variable (whichever way you wanna count).

Note that, though not necessary, it is also not entirely surprising that percentages and counts would go up with an item's count or that of its superset. The bottom graphs, on the other hand, are surprising, in so far as the item's count goes up along with that of its competitor. Such a correlation would not be sustained by *Huck Finn*, where many uncontracted competitors are completely excluded.

Furthermore, note that the correlation presented only includes a single kind of item. Krug does engage in some cross-item comparisons (*is/are*, p. 295), but it's not clear that the correlations are as consistent. While auxiliaries tend to be high frequency items, Krug's proposal seems to deny that they have any other particular affinity for contraction. Of course, their orthographic status is only marginal evidence of a particular affinity.

The tables in fig. 6, p. 20, give the top 100 bigrams with counts from our 2 novels, so that the reader may consider for themselves if more frequent bigrams are more collapsable (i.e. to their ear, rather than just in writing). Although these are only small subsets of English, their similarities and differences can give some perspective on the significance of string frequency. "Don't" and "didn't" (counted as a bigram) do appear very high in *Huck Finn*'s counts, but items like "was" and "the" presumably resist collapse because of semantic importance. Krug, in fact, recognizes this "recoverability principle" as a "counterforce" to his "Frequency Factor" (p. 310).

Thus, Krug's arguments might be more on track when suggesting, like Zipf, that high frequency is required to develop the close unions (as well as sustain the irregularities) of contraction. Like the shortness of high frequency words<sup>20</sup>, high frequency may be more of a necessity for contraction than a guarantee. However, even that kind of principle might have trouble accounting for contractions like "mightn't" without considering frequencies of sets of items.

<sup>20</sup>No high frequency words are long, but some low frequency words are short.

## 5 experiment proposal (acquisition)

However, in attempting to understand the input-driven component of learning, we should also hope that string frequency effects play some role in the acquisition process. Presumably, more common constructions should be easier to acquire. To test these assumptions, we can formulate the following experiment.

Using a database like CHILDES, we should be able to compare frequency counts in motherese with acquisition age in children's corpora. More frequent events should be acquired at a younger age. For instance, if the evidence in motherese sustained the frequency distinctions found in *Moby Dick*, then we would make the following predictions:

(20) predictions:

- a. Use of "don't" should precede "do not".
- b. Use of "did not" should precede "didn't".

If the input was like *Huck Finn*, on the other hand, we'd expect both contractions to be acquired first.

These predictions are based on both of the following hypotheses:

(21) hypotheses:

- a. Frequency facilitates acquisition.
- b. Decomposition is not more difficult than composition.<sup>21</sup>

Our string frequency hypothesis is qualified by a kind of relational hypothesis. If certain common constructions end up displaying acquisition difficulties, they may provide evidence of structural requirements<sup>22</sup> that are interfering with straight-forward, evidence-induced learning.

## 6 phylo≈onto

In seeking connections between variation and acquisition, we need to be wary of falling into Haeckel's wholesale equivalence.

- (22) "... ontogenesis is a brief and rapid recapitulation of phylogenesis, determined by the physiological functions of heredity (generation) and adaptation (maintenance)." -Haeckel, E. 1899. *Riddle of the Universe at the Close of the Nineteenth Century*.

Because of their distinct selectional pressures, where some sort of communicative standard is pre-established for individuals, Pinker and Bloom (1990) dismiss such connections.

---

<sup>21</sup>Note that the derivational feeding of current theory is somewhat invested in the notion that "unmoved" structures precede "moved" structures. If syntactic categories like *Aux* or *Neg* are innate, and strictly composed through Merger and Movement, then composition is necessary and decomposition would be impossible, except through some sort of reverse reasoning.

Also, different ratios of alternatives should probably affect acquisition rate.

<sup>22</sup>Such difficulties might also arise from developmental delays in general cognitive faculties.

(23) "... language evolution and language acquisition not only can differ but that they must differ." -Pinker and Bloom (1990), p. 719.

Yang (2003), on the other hand, because of his advocacy of a parameter-defined, finite hypothesis space, insists that species and individuals should explore the same possibility space.

(24) "If language is delimited in the finite space of Universal Grammar, its ontogeny might well recapitulate its scope and variations as the child gradually settles on one out of the many possibilities. This is exactly what the variational model leads one to expect, ..." -Yang (2003), p. 101.

However, even he would probably admit the individual's need to approach certain pre-existing standards could constrain his search more than the global variation of languages.

All the same, parallels can be sought in specific mechanisms. For instance, our interpretation of contraction has suggested that languages manifest, as learners acquire, bi-directional relations between coalesced and decomposed forms. Thus, evolution should be possible in both directions, and learners should be able to develop associations in both directions.

Different researchers have emphasized one process or the other. Arbib (2006) argues for a proto-language with singletons for relatively complex meanings such as "The alpha male has killed a meat animal and now the tribe has a chance to feast together. Yum, yum!" (p. 35); Hobbs (2006) (same book), on the other hand, suggests that decompositional reanalysis is 'not a very productive process, in contrast with "composition via discourse"' (p. 79). The acquisition of contraction alternatives should require that both mechanisms are productive.

In sum, though acquisition and variation face different pressures, there are processes that should be active factors in both, given that variation is ultimately, empirically dependent on the sum of acquisition processes that learners undergo. Coalescence and decomposition are claimed to be 2 such competing forces.

## 7 Goldvarb

Finally, a variable rule analysis was performed on the contraction data using Goldvarb. The Chi-square/cell value of less than 2 reflects a reasonable approximation of the model to the data.

Binomial, One Level:

- Chi-square/cell = 1.0995
- Log likelihood = -234.357

The stepping-up procedure prioritized the narrator effect (Voice: A=author, Q=quote), the specific auxiliary (Aux: D=do, C=can,could, T=shall, V=would, B=be, H=have), and preceding phonological context (Sonority: V=vowel, C=consonant), in that order, as shown by the span of weights in each factor group.

Binomial, Up and Down:

---

Run # 13, 26 cells:

Convergence at Iteration 9

Input 0.157

Group # 1 – V: 0.700, C: 0.416

Group # 2 – A: 0.189, Q: 0.856

Group # 6 – D: 0.744, C: 0.737, T: 0.462, V: 0.434, W: 0.549, B: 0.113, H: 0.496

Log likelihood = -243.664 Significance = 0.000

---

Adding order:

a. Voice (#2)

b. Aux (#6)

c. Sonority (#1)

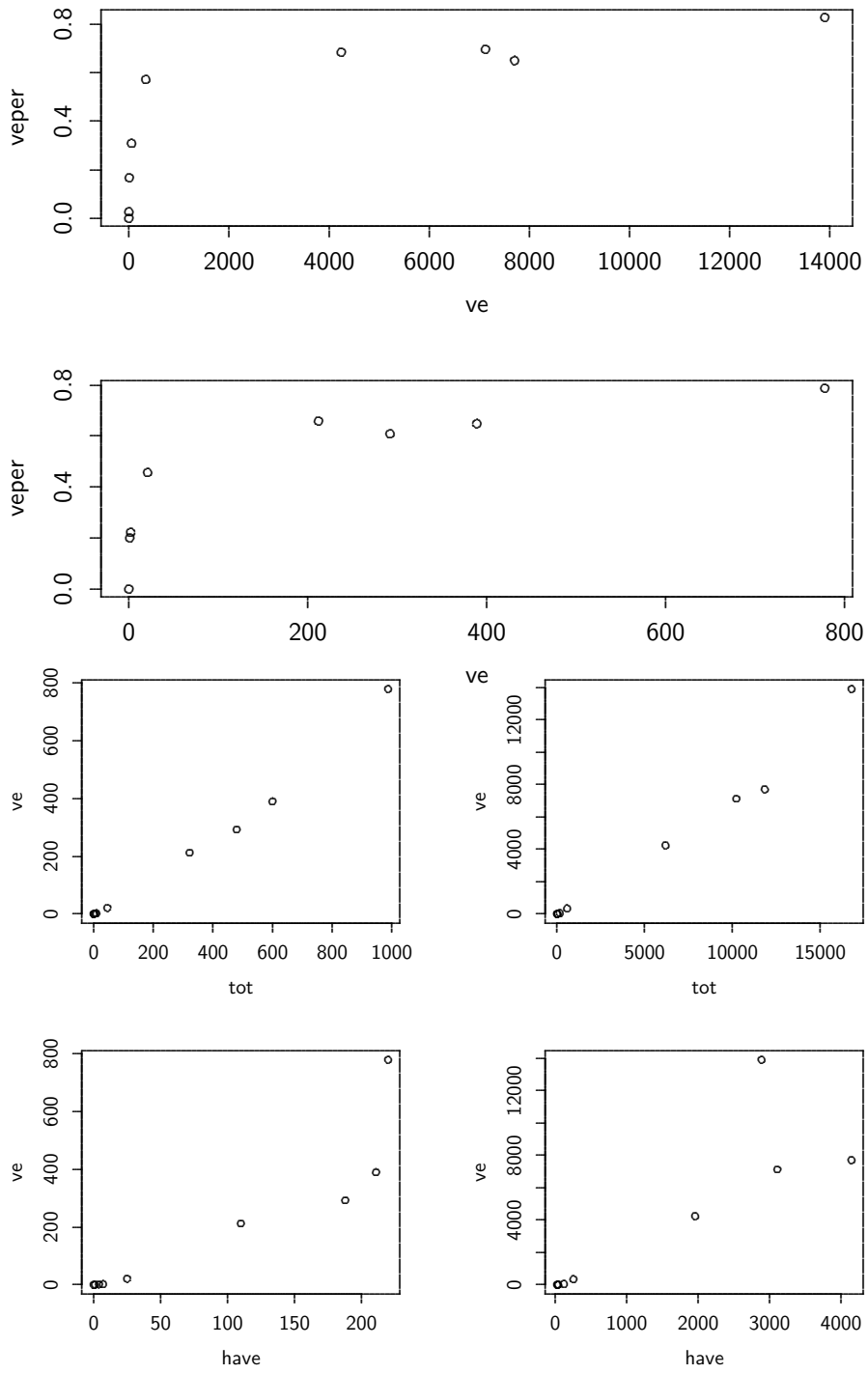
These results sustain the central claim that phonology and lexical specificity, as well as dialect, influence syntactic decisions.

## References

- Arbib, Michael A. (2006) "The Mirror System Hypothesis on the linkage of actions and languages," in Arbib, Michael A., ed., *Action to Language via the Mirror Neuron System*, pp. 3–47, Cambridge, Cambridge.
- Chomsky, Noam (1986) *Barriers*, MIT Press, Cambridge, Mass.
- Ellegård, Alvar (1953) *The Auxiliary Do*, Almquist & Wiksell (Göteborg), Stockholm.
- Eubank, Lynn and Grace, Sabine Thépaut (1998) "V-to-I and Inflection in Non-Native Grammars," in Beck, M.L., ed., *Morphology and its Interfaces in L2 Knowledge*, pp. 69–88, John Benjamins, Amsterdam.
- Hobbs, Jerry R. (2006) "The origin and evolution of language: a plausible, strong-AI account," in Arbib, Michael A., ed., *Action to Language via the Mirror Neuron System*, pp. 48–88, Cambridge, Cambridge.
- Krug, Manfred (1998) "String Frequency: A Cognitive Motivating Factor in Coalescence, Language Processing, and Linguistic Change," *Journal of English Linguistics*, 26(4), 286–320.
- Labov, William (1969) "Contraction, Deletion, and Inherent Variability of the English Copula," *Language*, 45(4), 715–762.
- Mohr, Eugene V. (1969) "The Independence of Contractions," *TESOL Quarterly*, 3(4), 285–289.
- Pinker, Steven and Bloom, Paul (1990) "Natural language and natural selection," *Behavioral and Brain Sciences*, 13(4), 707–784.
- Pollack, J-Y (1989) "Verb movement, universal grammar and the structure of IP," *Linguistic Inquiry*, 20, 365–424.
- Pullum, G.K. and Scholz, B.C. (2002) "Empirical assessment of stimulus poverty arguments," *The Linguistic Review*, 19, 9–50.

- Rizzi, Luigi (1990) *Relativized Minimality*, MIT Press, Cambridge, Mass.
- Roberts, Ian G. (1993) *Verbs and Diachronic Syntax*, Kluwer, Dordrecht.
- White, Lydia (2003) "Fossilization in steady state L2 grammars," *Bilingualism: Language and Cognition*, 6, 129–141.
- Yang, C.D. (1999) "A selectionist theory of language development," in *Proceedings of 37th Meeting of the Association for Computational Linguistics*, pp. 429–435, Association for Computational Linguistics, East Stroudsburg, PA.
- Yang, Charles D. (2000) "Internal and external forces in language change," *Language Variation and Change*, 12(3), 231–250.
- Yang, Charles D. (2003) *Knowledge and Learning in Natural Language*, Oxford, USA.
- Zwicky, Arnold M. (1970) "Auxiliary Reduction in English," *Linguistics Inquiry*, 1, 323–336.
- Zwicky, Arnold M. and Pullum, Geoffrey K. (1983) "Cliticization vs. Inflection: English N'T," *Language*, 59(3), 502–513.

Figure 5: Krug (1998):



Moby Dick		9121 s.		Huck Finn		6367 s.	
1835	of the	1142	in the	431	in the	369	it was
711	to the	432	from the	347	didn't	341	don't
380	the whale	367	of his	338	of the	317	and the
360	and the	340	on the	292	warn't	289	i was
326	at the	323	to be	288	ain't	287	and i
318	of a	314	by the	247	and then	234	to the
310	with the	302	for the	223	on the	220	was a
289	it was	279	it is	219	couldn't	214	it's
257	in his	254	in a	206	but i	203	he was
248	the ship	243	with a	200	out of	176	wouldn't
242	into the	220	upon the	171	and he	162	i says
218	the sea	213	as the	162	so i	157	in a
213	that the	198	all the	152	't no	145	going to
174	sperm whale	170	out of	142	that's	141	got to
168	for a	162	the same	141	the king	140	of a
157	the pequod	157	like a	138	to be	136	and so
148	one of	143	at last	134	a little	132	and a
142	he was	141	to his	131	and says	127	there was
140	the white	138	is the	125	for the	124	i didn't
136	but the	135	in this	124	the duke	123	all the
135	the other	135	and then	121	he said	120	i'll
134	over the	134	of all	118	to get	118	the river
134	such a	132	as if	113	to do	112	he says
130	of this	129	as a	112	and by	109	up and
129	of that	126	sort of	108	i couldn't	105	of it
123	was a	122	the boat	103	and it	102	of them
122	the sperm	120	whale's	102	kind of	101	all right
119	don't	119	have been	98	if i	97	can't
119	so that	114	the first	97	i got	97	i'd
114	he had	112	had been	96	to see	96	i could
111	the great	111	in that	95	's a	94	i see
111	to a	110	with his	94	when i	93	said he
110	is a	110	i have	91	at the	91	that was
107	a little	107	the most	91	the old	90	for a
107	and a	106	white whale	90	i reckon	90	a good
104	was the	103	his own	89	and we	89	't know
100	when the	99	it's	88	i had	88	by and
97	he's	96	the deck	88	and they	88	i don't
95	the captain	95	that he	87	on a	86	was all
95	and so	94	of it	86	up the	85	was the
94	that's	92	the old	85	done it	85	't you
92	to me	92	part of	84	want to	84	we was
92	through the	89	and all	84	then i	83	the way
89	and i	88	to him	83	then he	83	with a
88	a whale	88	and in	83	's the	82	one of
87	at all	87	but a	81	down the	80	and when
86	as he	85	the world	80	with the	79	he'd
84	from his	83	that is	79	i'm	79	she was
83	all this	81	moby dick	78	so he	77	the raft
81	old man	81	in all	77	we got	77	when he

Figure 6: top 100 bigrams