Getting Useful Gender Statistics from English Text

John Hale and Eugene Charniak

Department of Computer Science
Brown University
Providence, Rhode Island 02912

CS-98-06
May 1998

Getting Useful Gender Statistics from English Text

John Hale and Eugene Charniak
Department of Computer Science
Brown University

Abstract

Gender, understood as a lexical feature, is important for anaphora because it narrows down the number of possible referents involved in a typical pronoun resolution situation. This work describes an automatic method for obtaining reliable guesses about the gender of entities in a corpus using free text. By using a simple but unreliable anaphora algorithm repeatedly over a large corpus, the probable genders of referenced entities can be compiled and given a salience ranking. These statistics are an inexpensive way to add on gender-feature information to a statistical anaphora resolution algorithm.

Introduction

Anaphora is the use of a grammatical substitute, such as a pronoun, to refer to a word or group of words mentioned earlier in the discourse. To the degree that the referred-to word or group of words denotes some entity in the world, the anaphoric reference to that entity also refers to that same real-world object. By way of this correspondence, groups of words denoting the same individual, whether they are anaphoric or not, should share semantic features.

One semantic feature which undoubtedly exists on at least proper noun phrases is gender – whether the referenced entity is male, female or inanimate. In French, such a feature could be read off the morphological form of a noun, (perhaps using finite state techniques as in (Silberztein, 1987, page 98)) but in English virtually the only remaining reflex is in the form of referring pronouns[1]. Discourse entities continue to have gender in English, but the only time that gender becomes manifested is when the entity is referred to anaphorically with a gender-marked pronoun.

Furthermore, gender is a psychological factor that has been demonstrated (Ehrlich, 1980) to exert an influence on human anaphora resolution. Recent psychological theories

[1] Although in French we run into the distinction between "lexical" gender and "actual" gender, and issue which we will ignore in the following.

I am very grateful to the NLP group at Brown for help in this work, especially to Mark Johnson, Niyu Ge and Brian Roark.

(Gernsbacher, 1989) of referential access have posited gender to have an "enhancing" or "inhibiting" effect, suggesting a statistical characterization.

Here we present an automatic method for estimating the probability that nouns occurring in a large corpus of English text manifest masculine or feminine gender, or denote inanimate things. This method is based on simply counting co-occurrences of pronouns and noun phrases, and as such can employ any method of analysis of the text stream that results in referent/pronoun pairs (cf. (Hatzivassiloglou & McKeown, 1997) for another application in which no explicit indicators are available in the stream). We present two very simple methods for finding referent/pronoun pairs, and also present an application of a surprise statistic that can indicate how confident we should be about the predictions the method makes. Following this, we show the results of applying this method to the 21 million word 1987 Wall Street Journal corpus using two different putative pronoun reference strategies of varying sophistication, and evaluate their performance using word lists of names classified by gender.

Method

The method is a very simple mechanism for harvesting the kind of gender information present in discourse fragments like "Kim slept. She slept for a long time." Even if Kim's gender was unknown before seeing the first sentence, after the second sentence, it is known.

The probability that a referent is a particular gender class is just the relative frequency with which that referent is referred-to by a pronoun which is part of that gender class. That is, the probability of a referent *ref* being in gender class gc_i is

$$\Pr(ref \in gc_i) = \frac{\text{Count}(\text{references to } ref \text{ with pronoun } \in gc_i)}{\sum_{\text{all gender classes, } i} \text{Count}(\text{references to } ref \text{ with pronoun } gc_i)} \quad (1)$$

In this work we have only considered three gender classes: masculine, feminine and inanimate which are indicated by their typical pronouns, HE, SHE, and IT. However, a variety of pronouns indicate the same class: Plural pronouns like "they" and "us" don't

pronoun	*gender class*
he,himself,him,his	HE
she,herself,her, hers	SHE
it,itself,its	IT

reveal any gender information about their referent, and consequently aren't useful, although this might be a way to learn pluralization in an unsupervised manner.

Why is a statistical characterization of gender necessary, if the instant "snapping" to a definite gender occurs when disambiguating information comes along? The need for a statistical characterization derives from the fact that the actual referent of any given pronoun is not known – at least, not known by the program. Instead, a guess about the particular referent for a particular pronoun has to be used, in the hope that the guess will be correct often enough to produce a statistically salient pattern of observations. Because of this statistical approach to what is traditionally viewed as a discrete-valued semantic feature, this method can use unreliable pronoun reference guesses to learn useful features on corpus entities without depending on hand-coindexation of the corpus for error-free referent/pronoun co-occurrences.

Pronoun reference strategies

In order to gather statistics on the gender of referents in a corpus, there must be some way of identifying the referents. In attempting to bootstrap lexical information about referents' gender, we consider two strategies which are completely blind to any kind of semantics.

Previous Noun

One of the most naive pronoun reference strategies is the "previous noun" heuristic. On the intuition that referents in one sentence are often the objects of the previous sentence, this heuristic simply keeps track of the last noun seen, and then submits that noun as the referent of any pronouns following. This strategy is certainly simple-minded but it accounts for a healthy fraction of (43%) of pronoun reference situations in experiments performed by Niyu Ge on a hand-tagged subset of the Wall Street Journal corpus (see (Ge & Charniak, 1998)).

In the present system, a statistical parser using a grammar induced from held-out corpus is used (see (Charniak, 1997)) simply as a tagger. This apparent parser overkill is a control to ensure that the part of speech tags assigned to words are the same using the previous noun heuristic and the syntactic Hobbs algorithm, to be explained in the next section. In fact the only part of speech tags necessary are those indicating nouns and pronouns.

When this strategy is applied to the entire corpus, the result is a set of observations of pronoun/noun co-occurrences.

Syntactic Hobbs algorithm

Similarly, the Hobbs algorithm (Hobbs, 1976) also produces a set of observations of co-occurrences, although these are co-occurrences of full noun phrases with their referring pronoun. Unlike the previous noun heuristic, the Hobbs algorithm depends on having a parse of the current sentence (the one containing the pronoun to be resolved) available, as well as parses of the previous sentences in the text. The Hobbs algorithm is an attempt to make use of as many of the clues to pronoun reference as are available from a parse. To facilitate a comparison with the previous noun heuristic the whole output of the statistical parser is presented to the Hobbs algorithm, which in this implementation can fail to find a referent. For this reason there are fewer observations available to collect using the Hobbs algorithm. Nevertheless these observations are typically of much higher quality, both in the sense that they yield more plausible referents, and that these referents are correct more often. On the same hand-tagged subset of the corpus, Ms. Ge reports 65.3% accuracy with this strategy.

Surprise

Given a putative pronoun resolution method, and a corpus, the result is a set of pronoun/referent pairs. By collating by referent, and abstracting away to gender classes of pronouns, rather than individual pronouns, we have the relative frequencies with which a given referent is referred-to by pronouns of each gender class. The gender class for which this relative frequency is the highest we might say is the gender class to which the referent most probably belongs.

However, any syntax-only pronoun resolution strategy will be wrong some of the time – these methods know nothing about discourse boundaries, intensions or real-world knowledge. We would like to know, therefore, whether the pattern of pronoun references that we observe for a given referent is the result of our supposed "hypothesis about pronoun reference" – that is, the pronoun reference strategy we have provisionally adopted in order to gather statistics – or whether the pattern is the result of some other unidentified process.

This decision is made by ranking the referents by log-likelihood ratio, termed salience, for each referent with its associated pattern of references. The likelihood ratio is adapted from Dunning (Dunning, 1993, page 66) and uses the raw frequencies of each pronoun class in the corpus as the null hypothesis, $\Pr(gc0_i)$ as well as $\Pr(ref \in gc_i)$ from equation 1.

$$\text{salience} = -2\log\left(\frac{\prod_{\text{all classes } i} gc0_i^{observations\ with\ class\ i}}{\prod_{\text{all classes } i} \Pr(ref \in gc_i)^{observations\ with\ class\ i}}\right) \quad (2)$$

Making the unrealistic simplifying assumption that references of one gender class are completely independent of references via another classes[2], the likelihood function in this case is just the product over all classes of the probabilities of each class of reference to the power of the number of observations with this class.

Evaluation

Performing the salience computation over all of the referents, we are left with a list of referents that can be sorted by salience, and a prediction about what gender class each referent should fall into.

Qualitative

Often, for referents that are titles (i.e. "the President") or common nouns (i.e. "girl") it is straightforward to read off the gender class probabilities to determine if the prediction is correct. Because referents with higher salience scores are often explicitly-gendered common nouns like "mother" or "husband", this kind of evaluation yields a good qualitative indication of the technique's performance for those referents.

Quantitative evaluation for one subclass of referents

In comparing the total result of two initial pronoun resolution strategies, this subjective evaluation is inappropriate. Moreover, the qualitative evaluation can invoke world knowledge on the part of the evaluator unintentionally, as occurs when the evaluator knows the gender of some famous person or prominant business leader. A fully automatic evaluation can proceed only with some standardized knowledge of what proper names are masculine and which are feminine. To that end, we have devised a more objective test, useful only

[2] In effect, admitting that a referent could be in different gender classes across different observations.

for scoring the subset of referents that are names of people. This scoring uses publically available word lists of first names that come separated into male (MASC)and female (FEM) lists[3]. They have been augmented with the honorifics "Mr." "Mrs." and "Ms." Should the referent be present in the list corresponding to the gender class prediction, that is a correct classification. In the case of the Hobbs algorithm, which returns full noun phrases as referents, inclusion of any word from the noun phrase in the appropriate list constitutes a correct classification; this means "Mr. John Smith" is counted as correct if the statistics indicate that it is in the masculine gender class, (on account of "John" being in the masculine list) but it also means that "Jean Marie Le Pen" would not be counted as correct if "Marie" fails to appear in the masculine list but does appear in the feminine list (assuming none of the other words appear in either list). This score is the precision of gender attribution for names containing words in the first name lists:

$$\text{precision} = \frac{\begin{array}{c}\#\text{ times } \mathit{ref} \text{ attributed as } \textsc{he} \wedge \text{contains a word in } \textsc{masc} \ +\\ \#\text{ times } \mathit{ref} \text{ attributed as } \textsc{she} \wedge \text{contains a word in } \textsc{fem}\end{array}}{\#\text{ times } \mathit{ref} \text{ contains a word in either list}} \tag{3}$$

Results

Results of this gender-determination technique carried out on the 1987 Linguistic Data Consortium Wall Street Journal corpus are given in the appendices, which include the top and bottom of the ranked salience list. The format of the results is as follows: The referent is given, followed by the total number of references by pronouns of all types, with the salience in parenthesis. Then the probabilities and individual counts of reference observations for each class are given. Class 0 is HE-type, 1 is SHE-type and 2 is IT-type.

Qualitative impressions

Qualitatively speaking, results with salience values above about 100 seem to square with our intuitions about the genders of people mentioned in the Wall Street Journal, or are easily verified correct by searching for the referent; for example MOODY is predicted to be IT-type, and in fact it is a proper name – but a proper name of a company, Moody's Investors Service Inc. Saliences below 100 contain many correct classifications, but are not uniformly correct.

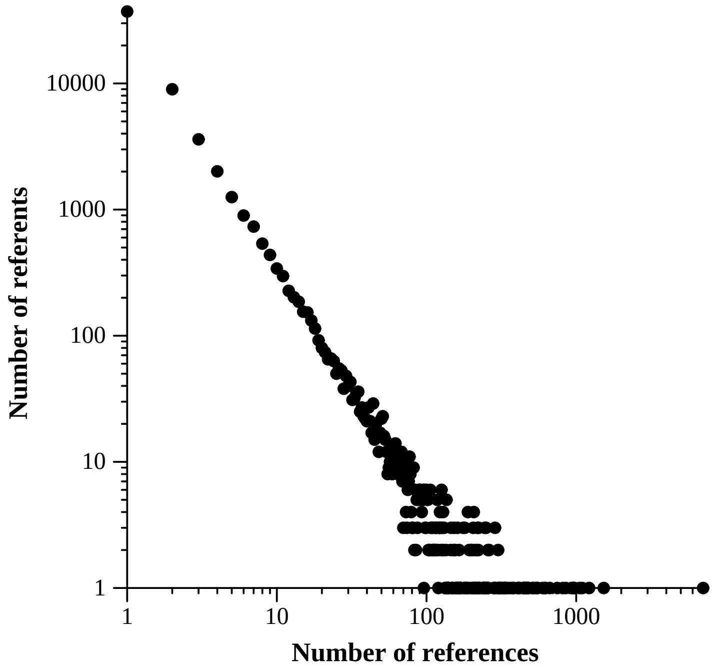
Precision scores

The precision score computed over all names from either name list appearing in referent are 38.8% for the previous noun heuristic and 68.15% for the syntactic Hobbs algorithm.

These scores correspond very closely with the experimental results of Niyu Ge applying each strategy to smaller amounts of manually parsed and coindexed text. On one hand, we should expect such an automatic method as this, relying as it does on automatically-generated part-of-speech tags and parses from an automatically induced grammar to perform worse, even on the restricted subset of peoples' names as referents. On the other, we might expect it to perform better, given that it has much greater amounts of data over

[3]These lists are courtesy of Paul Leyland and are available at `<ftp://ftp.ox.ac.uk/pub/wordlists/>`

Figure 1. Pronoun references obey Zipf's law



Zipfian pronoun reference using Syntactic Hobbs algorithm on WSJ 1987

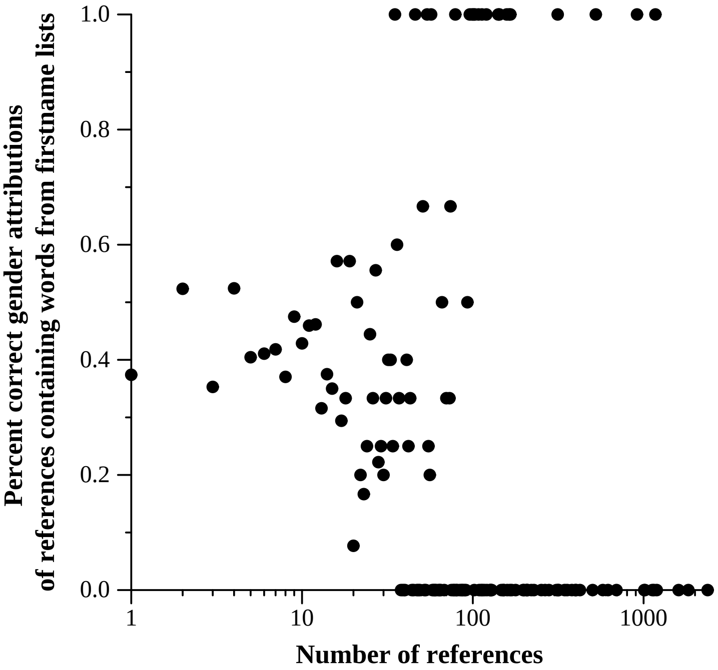
which errors might in some sense be amortized. The close numerical agreement (43% with 38.8% and 65.3% with 68.15%) indicates that these two factors roughly cancel. In the case of the Hobbs-based figure, there appears to be no appreciable loss of accuracy using a statistical parser as compared with using the gold standard parses.

Though there is more data in general, however, useful levels of reference data for a particular referent are rare. This tradeoff can be seen in the Zipfian relationship between number of references to a referent, and the number of such referents that we find in the corpus.

Precision graphs

Just as a higher salience implies a higher likelihood of being correctly classified, we would expect higher numbers of total references for a given referent to be associated with higher likelihoods of being correct. Graphs 2 and 3 do not bear out this expectation. Rather, for almost every referent with more than about 50 recorded references, there is a trimodal pattern of precision. This is result which illustrates how gender attribution is really a special case of word-sense disambiguation. About fifty references is all it takes for this gender-attribution method to "be sure" about its attribution for a given referent. In many of these cases the attribution is to the inanimate or IT-type gender class. The evaluation metric, however, noting the presence of the referent, or a constituent word of the referent in one of the first name lists assumes that the correct gender is one of the animate ones, and marks the attribution wrong, when in fact the attribution was correct, but only if the referent is interpreted as an inanimate object and not the name of a person. This happens

Figure 2. Precison using first name scoring scheme with previous noun heuristic



Percentage correct on proper names
in gender-classified first names using previous-noun heuristic on WSJ 1987

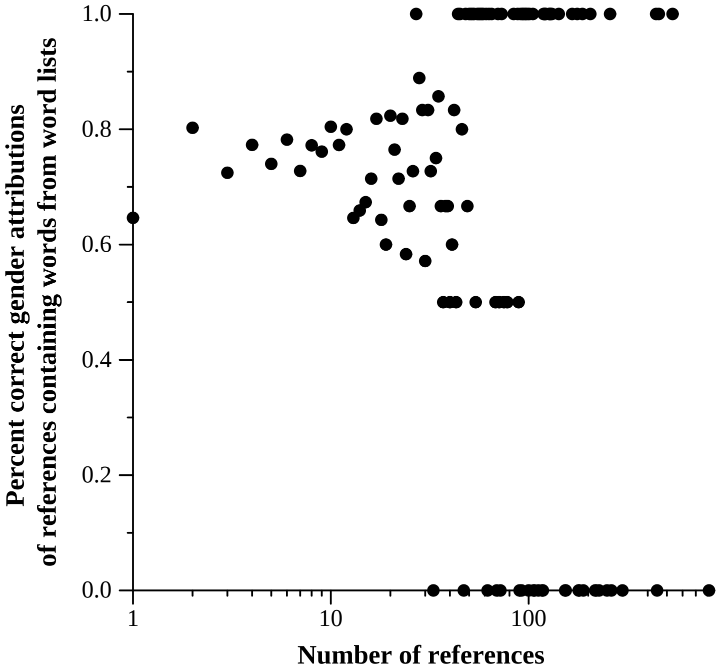
in cases such as "the Reagan Administration" which the first name-based evaluation metric marks as masculine since "Reagan" is present in the MASC list. Further examples of this limitation of the scoring metric include names like "Bank," "Way" (short for Wayland), "Case" (short for Casey), "Court" (short for Courtney), "Ford," "Law" and "Price."

We are currently investigating other evaluation methods, using unambiguous honorifics (only), disjunctions of name lists and machine-readable dictionaries, and name phonology.

Conclusions and Future Directions

Gender information does not have to be read off a dictionary. Conceived of as just another semantic feature, there are a variety of textual analyses that yielding observations of pronouns/referent pairs which will result in reasonable gender attributions for entities in a corpus. While English gender features are inferentially deeper than in French, in many cases they can be inferred statistically from large corpora, with no human intervention. Unsurprisingly, more sophisticated pronoun reference strategies produce better results than simpler ones, and we can be more confident about cases in which there is a lot of data than those in which data is sparse. Because these cases can be distinguished, however, we can define the level of confidence we need and specify the amount of corpora needed accordingly. An interesting followup might be to relate this amount to the amount of speech input apprehended by children, and to relate the sophistication of children's pronoun reference

Figure 3. Precision using first name scoring scheme with syntactic Hobbs algorithm



Percentage correct on proper names
in gender-classified first names using syntactic Hobbs algorithm on WSJ 1987

strategies to those used in this statistical method.

Future work that has been initiated is examining the possibility of feeding back the gender information gathered into a pronoun reference algorithm that does know about semantics, and a little about discourse, both in a statistical form.

References

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the 14th National Conference on Artificial Intelligence* Menlo Park, CA. AAAI Press/MIT Press.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1).

Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology*, *32*, 247–255.

Ge, N. & Charniak, E. (1998). A statistical approach to anaphora resolution. Forthcoming.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99–156.

Hatzivassiloglou, V. & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives..

Hobbs, J. R. (1976). Pronoun resolution. Tech. rep., City College of New York.

Silberztein, M. (1987). The lexical analysis of french. In M. Gross & D. Perrin (Eds.), *Electronic Dictionaries and Automata in Computational Linguistics*, Vol. 377 of *Lecture Notes in Computer Science* Saint-Pierre d'Oléron, France. LITP Spring School on Theoretical Computer Science.

Previous Noun Heuristic

```
there were 514015 pronouns seen overall
size of pairing class 0: 15059
size of pairing class 1: 4884
size of pairing class 2: 21346
total dictionary is 26627 words
null references was 0
total for category 0 is 129015
prob of 0 is 0.250995
total for category 1 is 16043
prob of 1 is 0.0312112
total for category 2 is 368957
prob of 2 is 0.717794
CORP. , refs=5314(796.948) 0 - 0.0538201 refs=286 1 - 0.00131728 refs=7 2 - 0.944863 refs=5021
INC. , refs=5472(609.776) 0 - 0.0752924 refs=412 1 - 0.00493421 refs=27 2 - 0.919773 refs=5033
COMPANY , refs=9979(520.038) 0 - 0.128871 refs=1286 1 - 0.0096202 refs=96 2 - 0.861509 refs=8597
REAGAN , refs=1173(416.198) 0 - 0.681159 refs=799 1 - 0.0179028 refs=21 2 - 0.300938 refs=353 MASC CORRECT

WOMAN , refs=403(378.088) 0 - 0.176179 refs=71 1 - 0.503722 refs=203 2 - 0.320099 refs=129
PRESIDENT , refs=1997(341.819) 0 - 0.545819 refs=1090 1 - 0.016024 refs=32 2 - 0.438157 refs=875
% , refs=5419(328.15) 0 - 0.121978 refs=661 1 - 0.0071969 refs=39 2 - 0.870825 refs=4719
MAN , refs=915(247.551) 0 - 0.615301 refs=563 1 - 0.042623 refs=39 2 - 0.342077 refs=313 MASC CORRECT

NORTH , refs=525(187.933) 0 - 0.681905 refs=358 1 - 0.0228571 refs=12 2 - 0.295238 refs=155 MASC CORRECT

CO. , refs=2989(185.641) 0 - 0.113081 refs=338 1 - 0.0120442 refs=36 2 - 0.874875 refs=2615
U.S. , refs=3368(157.335) 0 - 0.135689 refs=457 1 - 0.010095 refs=34 2 - 0.854216 refs=2877
AQUINO , refs=114(149.784) 0 - 0.122807 refs=14 1 - 0.622807 refs=71 2 - 0.254386 refs=29
MOTHER , refs=243(131.208) 0 - 0.26749 refs=65 1 - 0.349794 refs=85 2 - 0.382716 refs=93
CONCERN , refs=1843(123.067) 0 - 0.116658 refs=215 1 - 0.00596853 refs=11 2 - 0.877374 refs=1617
CHAIRMAN , refs=704(119.165) 0 - 0.544034 refs=383 1 - 0.015625 refs=11 2 - 0.440341 refs=310
GORBACHEV , refs=338(118.67) 0 - 0.671598 refs=227 1 -   0 refs=0 2 - 0.328402 refs=111
SALE , refs=1274(117.952) 0 - 0.0918367 refs=117 1 - 0.00470958 refs=6 2 - 0.903454 refs=1151
THATCHER , refs=93(115.384) 0 - 0.0967742 refs=9 1 - 0.602151 refs=56 2 - 0.301075 refs=28
GROUP , refs=2263(109.904) 0 - 0.12859 refs=291 1 - 0.0132567 refs=30 2 - 0.858153 refs=1942
GOVERNMENT , refs=2777(106.75) 0 - 0.148001 refs=411 1 - 0.0104429 refs=29 2 - 0.841556 refs=2337


...27269 lines deleted...

GRIDIRON , refs=9(-1.82054) 0 - 0.111111 refs=1 1 -   0 refs=0 2 - 0.555556 refs=5
PARTY , refs=672(-2.09004) 0 - 0.270833 refs=182 1 - 0.0223214 refs=15 2 - 0.700893 refs=471
CALAF , refs=13(-2.9671) 0 - 0.0769231 refs=1 1 -   0 refs=0 2 - 0.0769231 refs=1
CHANG , refs=18(-3.91963) 0 - 0.222222 refs=4 1 - 0.0555556 refs=1 2 - 0.333333 refs=6
TRAIL , refs=51(-7.46297) 0 - 0.176471 refs=9 1 - 0.117647 refs=6 2 - 0.294118 refs=15
SHEEP , refs=31(-8.79631) 0 - 0.0645161 refs=2 1 -   0 refs=0 2 - 0.387097 refs=12
Score --------------------------------------------------
 masc_correct = 585
 fem_correct = 84
 total list hits (correct or not) = 1680
 ((masc_correct + fem_correct) / total list hits) = 0.398214
```

Syntactic Hobbs Algorithm

```
Total number of ticks: 329652
size of pairing class 0: 29477
size of pairing class 1: 5413
size of pairing class 2: 36303
total dictionary is 59267 words
null references was 75072
total for category 0 is 92590
prob of 0 is 0.363697
total for category 1 is 10439
prob of 1 is 0.0410048
total for category 2 is 151551
prob of 2 is 0.595298
COMPANY  , refs=7052(1629.39) 0 - 0.0764322 refs=539 1 - 0.00609756 refs=43 2 - 0.91747 refs=6470
WOMAN  , refs=250(368.267) 0 - 0.172 refs=43 1 - 0.708 refs=177 2 - 0.12 refs=30
PRESIDENT  , refs=931(356.539) 0 - 0.820623 refs=764 1 - 0.0139635 refs=13 2 - 0.165414 refs=154
GROUP  , refs=1096(287.319) 0 - 0.060219 refs=66 1 - 0.00547445 refs=6 2 - 0.934307 refs=1024
MR. REAGAN  , refs=534(270.8) 0 - 0.882022 refs=471 1 - 0.00374532 refs=2 2 - 0.114232 refs=61 MASC CORRECT

MAN  , refs=441(202.102) 0 - 0.848073 refs=374 1 - 0.0385488 refs=17 2 - 0.113379 refs=50 MASC CORRECT

PRESIDENT REAGAN  , refs=455(194.928) 0 - 0.843956 refs=384 1 - 0.0043956 refs=2 2 - 0.151648 refs=69 MASC CORRECT

GOVERNMENT  , refs=1220(194.187) 0 - 0.117213 refs=143 1 - 0.0122951 refs=15 2 - 0.870492 refs=1062
U.S.  , refs=969(188.468) 0 - 0.102167 refs=99 1 - 0.00412797 refs=4 2 - 0.893705 refs=866
BANK  , refs=816(161.23) 0 - 0.0955882 refs=78 1 - 0.00735294 refs=6 2 - 0.897059 refs=732
MOTHER  , refs=113(161.204) 0 - 0.300885 refs=34 1 - 0.654867 refs=74 2 - 0.0442478 refs=5
COL. NORTH  , refs=258(158.692) 0 - 0.926357 refs=239 1 - 0.00775194 refs=2 2 - 0.0658915 refs=17 MASC CORRECT

MOODY  , refs=383(152.405) 0 - 0.0078329 refs=3 1 - 0.00522193 refs=2 2 - 0.986945 refs=378
SPOKESWOMAN  , refs=139(145.627) 0 - 0.122302 refs=17 1 - 0.582734 refs=81 2 - 0.294964 refs=41
MRS. AQUINO  , refs=73(142.223) 0 - 0.0958904 refs=7 1 - 0.835616 refs=61 2 - 0.0684932 refs=5
MRS. THATCHER  , refs=68(128.306) 0 - 0.0735294 refs=5 1 - 0.823529 refs=56 2 - 0.102941 refs=7
GM  , refs=513(119.664) 0 - 0.0779727 refs=40 1 - 0.00389864 refs=2 2 - 0.918129 refs=471
PLAN  , refs=514(111.134) 0 - 0.0856031 refs=44 1 - 0.00583658 refs=3 2 - 0.90856 refs=467
MR. GORBACHEV  , refs=205(108.776) 0 - 0.892683 refs=183 1 - 0.00487805 refs=1 2 - 0.102439 refs=21
JUDGE BORK  , refs=212(108.746) 0 - 0.882075 refs=187 1 -   0 refs=0 2 - 0.117925 refs=25
HUSBAND  , refs=91(107.438) 0 - 0.362637 refs=33 1 - 0.571429 refs=52 2 - 0.0659341 refs=6
JAPAN  , refs=450(100.727) 0 - 0.0755556 refs=34 1 - 0.0111111 refs=5 2 - 0.913333 refs=411
AGENCY  , refs=476(97.4016) 0 - 0.0840336 refs=40 1 - 0.0147059 refs=7 2 - 0.901261 refs=429
WIFE  , refs=153(93.7485) 0 - 0.614379 refs=94 1 - 0.287582 refs=44 2 - 0.0980392 refs=15
DOLLAR  , refs=621(90.8963) 0 - 0.130435 refs=81 1 - 0.00966184 refs=6 2 - 0.859903 refs=534
STANDARD POOR  , refs=200(90.1062) 0 -   0 refs=0 1 -   0 refs=0 2 -   1 refs=200
FATHER  , refs=146(89.4178) 0 - 0.808219 refs=118 1 - 0.143836 refs=21 2 - 0.0479452 refs=7
UTILITY  , refs=242(87.1821) 0 - 0.0247934 refs=6 1 -   0 refs=0 2 - 0.975207 refs=236
MR. TRUMP  , refs=129(86.5345) 0 - 0.945736 refs=122 1 - 0.00775194 refs=1 2 - 0.0465116 refs=6
MR. BAKER  , refs=187(84.2796) 0 - 0.855615 refs=160 1 - 0.00534759 refs=1 2 - 0.139037 refs=26
IBM  , refs=316(82.4361) 0 - 0.0696203 refs=22 1 -   0 refs=0 2 - 0.93038 refs=294
MAKER  , refs=224(82.252) 0 - 0.0223214 refs=5 1 -   0 refs=0 2 - 0.977679 refs=219
YEARS  , refs=1055(82.1632) 0 - 0.529858 refs=559 1 - 0.0815166 refs=86 2 - 0.388626 refs=410
MR. MEESE  , refs=166(82.1007) 0 - 0.873494 refs=145 1 -   0 refs=0 2 - 0.126506 refs=21
BRAZIL  , refs=285(79.7311) 0 - 0.0596491 refs=17 1 -   0 refs=0 2 - 0.940351 refs=268
SPOKESMAN  , refs=665(78.3441) 0 - 0.607519 refs=404 1 - 0.00451128 refs=3 2 - 0.38797 refs=258
%  , refs=851(76.9956) 0 - 0.185664 refs=158 1 - 0.00940071 refs=8 2 - 0.804935 refs=685
MR. SIMON  , refs=105(72.6446) 0 - 0.952381 refs=100 1 -   0 refs=0 2 - 0.047619 refs=5 MASC CORRECT

DAUGHTER  , refs=47(71.3863) 0 - 0.234043 refs=11 1 - 0.702128 refs=33 2 - 0.0638298 refs=3
FORD  , refs=249(71.3603) 0 - 0.0562249 refs=14 1 -   0 refs=0 2 - 0.943775 refs=235
MR. GREENSPAN  , refs=120(68.7807) 0 - 0.908333 refs=109 1 -   0 refs=0 2 - 0.0916667 refs=11
AT&T  , refs=198(67.9668) 0 - 0.0252525 refs=5 1 - 0.00505051 refs=1 2 - 0.969697 refs=192
MINISTER  , refs=125(67.7475) 0 - 0.864 refs=108 1 - 0.064 refs=8 2 - 0.072 refs=9
JUDGE  , refs=239(67.5899) 0 - 0.715481 refs=171 1 - 0.083682 refs=20 2 - 0.200837 refs=48


...75885 lines deleted...


STATUS  , refs=22(0.0047653) 0 - 0.363636 refs=8 1 - 0.0454545 refs=1 2 - 0.590909 refs=13
CASH FLOW  , refs=25(0.00108112) 0 - 0.36 refs=9 1 - 0.04 refs=1 2 - 0.6 refs=15
TURMOIL  , refs=25(0.00108112) 0 - 0.36 refs=9 1 - 0.04 refs=1 2 - 0.6 refs=15
GATE  , refs=7(-0.561412) 0 - 0.285714 refs=2 1 -   0 refs=0 2 - 0.571429 refs=4
MASS.  , refs=4(-1.07919) 0 - 0.25 refs=1 1 -   0 refs=0 2 - 0.25 refs=1
GUILD  , refs=15(-2.23743) 0 - 0.0666667 refs=1 1 -   0 refs=0 2 - 0.533333 refs=8
NOTE  , refs=51(-13.1074) 0 - 0.294118 refs=15 1 - 0.0196078 refs=1 2 - 0.215686 refs=11
Score --------------------------------------------------
 masc_correct = 10357
 fem_correct = 907
 total list hits (correct or not) = 18292
 ((masc_correct + fem_correct) / total list hits) = 0.615788
```