Bayesian Identification of Cognates and Correspondences

T. Mark Ellison
Linguistics, University of Western Australia,
and Analith Ltd
mark@markellison.net

Abstract

This paper presents a Bayesian approach to comparing languages: identifying cognates and the regular correspondences that compose them. A simple model of language is extended to include these notions in an account of parent languages. An expression is developed for the posterior probability of child language forms given a parent language. Bayes' Theorem offers a schema for evaluating choices of cognates and correspondences to explain semantically matched data. An implementation optimising this value with gradient descent is shown to distinguish cognates from non-cognates in data from Polish and Russian.

Modern historical linguistics addresses questions like the following. How did language originate? What were historically-recorded languages like? How related are languages? What were the ancestors of modern languages like? Recently, computation has become a key tool in addressing such questions.

Kirby (2002) gives an overview of current current work on how language evolved, much of it based on computational models and simulations. Ellison (1992) presents a linguistically motivated method for classifying consonants as consonants or vowels. An unexpected result for the dead language Gothic provides added weight to one of two competing phonological interpretations of the orthography of this dead language.

Other recent work has applied computational methods for phylogenetics to measuring linguistic distances, and/or constructing taxonomic trees from distances between languages and dialects (Dyen et al., 1992; Ringe et al., 2002; Gray and Atkinson, 2003; McMahon and McMahon, 2003; Nakleh et al., 2005; Ellison and Kirby, 2006).

A central focus of historical linguistics is the reconstruction of parent languages from the evidence of their descendents. In historical linguistics proper, this is done by the comparative method (Jeffers and Lehiste, 1989; Hock, 1991) in which shared arbitrary structure is assumed to reflect common origin. At the phonological level, reconstruction identifies cognates and correspondences, and then constructs sound changes which explain them.

This paper presents a Bayesian approach to assessing cognates and correspondences. Best sets of cognates and correspondences can then be identified by gradient ascent on this evaluation measure. While the work is motivated by the eventual goal of offering software solutions to historical linguistics, it also hopes to show that Bayes' theorem applied to an explicit, simple model of language can lead to a principled and tractable method for identifying cognates.

The structure of the paper is as follows. The next section details the notions of historical linguistics needed for this paper. Section 2 formally defines a model of language and parent language. The subsequent section situates the work amongst similar work in the literature,

Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology, pages 15–22,
Prague, June 2007. ©2007 Association for Computational Linguistics

making use of concepts described in the earlier sections. Section 4 describes the calculation of the probability of wordlist data given a hypothesised parent language. This is combined with Bayes' theorem and gradient search in an algorithm to find the best parent language for the data. Section 5 describes the results of applying an implementation of the algorithm to data from Polish and Russian. The final section summarises the paper and suggests further work.

1 Cognates, Correspondences and Reconstruction

In the neo-Grammarian model of language change, a population speaking a uniform language divides, and then the two populations undergo separate language changes.

Word forms with continuous histories in respective daughter languages descending which from a common word-form ancestor are called **cognate**, no matter what has happened to their semantics. Cognate word forms may have undergone deformations to make them less similar to each other, these deformations resulting from regular, phonological changes. Note that in the fields of applied linguistics, second language acquisition, and machine translation, the term *cognate* is used to mean any words that are phonologically similar to each other. This is not the sense meant here.

Phonological change produces modifications to the segmental inventory, replacing one segment by another in all or only some contexts. This sometimes has the effect of collapsing segment types together. Other changes may divide one segment type into two, depending on a contextual condition. The relation of parent-language segments to daughter-language segments is, usually, a many-to-many relation.

Parent-child segmental relations are reflected in the correspondences between segment inventories in the daughter languages. **Correspondences** are pairings of segments from daughter languages which have derived from a common parent segment. For example, **p** in Latin frequently corresponds to **f** in English, as in words like **pater** and **father**. Both segments have developed from a (postulated) Proto-IndoEuropean ***p**. Because correspondences only occur between cognates, identifying the two is often a bootstrap process: corraling cognates helps find more correspondences, and forms sharing a number correspondences are probably cognate.

2 Formal Structures

The method presented in this paper is based on a formal model of language. This is described in section 2.1. The subsequent section extends the model to define a parent language, whose segmental inventory is correspondences and whose lexicon is cognates linking two descendent languages.

2.1 Language model

The language model is based on three assumptions.

Assumption 1 *There is a universal, discrete set M of meanings.*

Assumption 2 *A language L has its own set of segments $\Sigma(L)$.*

Assumption 3 *The lexicon λ of a language L is a partial map of meanings to strings of segments $\lambda : M \to \Sigma(L)^*$.*

On the basis of these assumptions, we can define a language L to be a triple $(M, \Sigma(L), \lambda(L))$ of meanings, segments and mappings from meanings onto strings of segments.

For example, consider written Polish. The set of meanings contains concepts as TO TAKE-perfect-infinitive, TREE-nominative-singular, and so on. The segmental inventory contains the 32 segments **a ą b c ć d e ę f g h i j k l ł m n ń o ó p r s ś t u w y z ż ź**, ignoring capitalisation. The lexicon matches meanings to strings of segments, TO TAKE-perfect-infinitive to **wziąć**, TREE-nominative-singular to **drzewo**.

2.2 Parent language model

Definition 1 *A degree-(u, v) **correspondence** between L_1 and L_2 is a pair of strings $(s, t) \in \Sigma(L_1) \times \Sigma(L_2)$ over the segments of L_1 and L_2*

respectively, with lengths at least u and no more than v.

As an example of a correspondence, consider the pair of small strings from Polish and Russian, **(ć,ть)**. This is a degree-$(1,2)$ correspondence because its members have lengths as low as one and as high as two. It is also a degree-(u,v) correspondence for any $u \leq 1$ and $v \geq 2$.

Any correspondence can be mapped onto its components by projection functions.

Definition 2 *The **projections** π_1 and π_2 map a correspondence (s,t) onto its first $\pi_1(s,t) = s$ or second $\pi_2(s,t) = t$ component string respectively.*

The first projection function will map **(ć,ть)** onto **ć**, while the second maps **(ć,ть)** onto **ть**.

Correspondences can be formed into strings. These strings also have projections.

Definition 3 *The **projections** π_1 and π_2 map a string of correspondences $c_1..c_k$ onto the concatenation of the projections of each correspondence.*

$$\pi_1(c_1..c_k) = \pi_1(c_1)\pi_1(c_2)..\pi_1(c_k),$$

$$\pi_2(c_1..c_k) = \pi_2(c_1)\pi_2(c_2)..\pi_2(c_k)$$

Suppose we sequence four correspondences into the string **(w,в)(z,з)(ią,я)(ć,ть)**. This string has first and second projections, **wziąć** and **взять**, formed by concatenating the respective projections of each correspondence.

We can now define a parent language.

Definition 4 *A degree-(u,v) **parent** L_0 of two languages L_1, L_2 is a triple $(M, \Sigma(L_0), \lambda(L_0))$ where $\Sigma(L_0)$ is a set of degree-(u,v) correspondences between L_1 and L_2, excluding the pair of null strings, and $\lambda(L_0)$ is a partial mapping from M onto $\Sigma(L_0)$ which obeys*

$$\pi_1 \circ \lambda(L_0) \subseteq \lambda(L_1),\ \pi_2 \circ \lambda(L_0) \subseteq \lambda(L_2)$$

The circle stands for function composition.

Continuing our past example, we will focus on the two meanings TO TAKE-perfect-infinitive and TREE-nominative-singular. The segment inventory for the parent language contains degree-$(0,2)$ correspondences: **(,e)**, **(ć,ть)**, **(d,д)**, **(e,e)**, **(ią,я)**, **(o,o)**, **(rz,р)**, **(w,в)**, **(z,з)**. The lexical function maps TO TAKE-perfect-infinitive onto the string of correspondences **(w,в) (z,з) (ią,я) (ć,ть)** while TREE-nominative-singular maps to **(d,д) (,e) (rz,р) (e,e) (w,в) (o,o)**.

The parent language condition is verified by checking the projections of the two correspondence strings. The first string has projections **wziąć** and **взять**, which are forms for the meaning TO TAKE-perfect-infinitive in Polish and Russian respectively. The second string has projections **drzewo** and **дерево**, which are forms for the meaning TREE-nominative-singular in Polish and Russian respectively. So the projection condition is satisfied. If the lexical function is only defined on these two meanings, then this is a valid parent language.

It is worth emphasising that the projection condition for qualifying as a parent language applies only for those meanings for which the parent lexical mapping is defined. The corresponding forms in the child languages are said to be **cognate** in this model. Where no parent form is reconstructed, the forms are not cognate, and are to be accounted for in some way other than the parent language.

3 Related Work

The current work is, of course, far from the first to seek to identify cognates and/or correspondences. Here is an abbreviated overview of previous work in the field[1]. More detailed surveys can be found in chapter 3 of Kondrak's (2002) PhD thesis or Lowe's online survey [2] of prior art in this field.

In perhaps the first computational work on historical linguistics, Kay (1964) described an algorithm for determining correspondences given a list of cognate pairs across two daughter languages. His method seeks to find the smallest set

[1] An anonymous reviewer suggests that the current work shares features with that of Kessler (2001). I have been unable to access this book in time to include discussion of it in this paper.

[2] linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html

of correspondences which allows a degree-$(1, \infty)$ alignment for each cognate pair. Unfortunately, the complexity of the problem has precluded its application to significant daa sets.

Frantz (1970) developed a PL/1 programming which returned numerical evaluations of correspondences and cognacy, given a list of possible cognate word-pairs. Each word pair must be supplied as a degree-$(0, 1)$ reconstruction, that is, aligning single segments with each other or with gaps.

Guy (1984; 1994) presented a program called COGNATE which finds regular correspondences and identifies cognates using statistical techniques.

For his Master's, Broza (1998) developed MDL-based software called *candid* which identifies correspondences from cognates and expresses these as contextual phonological transformation rules.

Kondrak's (2002) doctoral dissertation combines phonological and semantic similarity methods with correspondance-learning. The algorithms for learning correspondences are taken from Melamed's (2000) probabilistic methods for identifying word-word translation equivalence. These methods, like the current work, are Bayesian. Because Melamed's problem seeks partial rather than complete explanation of the inputs in terms of correspondences, the matching problem is somewhat more difficult theoretically. As a result, he does not arrive at the decomposition of the sum of the probability of two inputs given the set of possible correspondences, approximating this with a high probability alignment.

4 Conditional Probability of the Data

The core of any Bayesian model is the conditional probability of the data given the hypothesis. This section details how probabilities assigned to data, and the assumptions on which this assignment is based.

The data is the mapping of meanings onto forms in two daughter languages. If those two languages are L_1 and L_2, we want to determine $P(\lambda(L_1), \lambda(L_2)|h)$. The nature of h will be discussed in section 4.6.

For brevity, we will write λ_i for $\lambda(L_i)$.

4.1 Meaning independence

The first step in defining the conditional probability of the data is to decompose it into meaning-by-meaning probabilities. This can be achieved by adopting the following two assumptions.

Assumption 4 *In a given language, the forms for different meanings are selected independently.*

This assumption states that *within a single language* choosing, for example, a form **wziąć** for meaning TO TAKE-perfect-infinitive is no help in predicting the form which expresses TREE-nominative-singular.

Assumption 5 *Across different languages, the forms corresponding to different meanings are independent.*

According to this assumption, the Polish word **wziąć** and the Russian word **взять** can be structurally dependent because they express the same meaning. In contrast, we can only expect a chance relationship between the Russian word **взять** meaning TO TAKE-perfect-infinitive, and the Polish word **drzewo** expressing TREE-nominative-singular.

Together, these two assumptions imply that the only dependencies possible between any four forms expressing the two meanings m_1 and m_2 in two languages L_1 and L_2 are between $\lambda(m_1)$ and $\lambda(m_1)$ on the one hand and $\lambda(m_2)$ and $\lambda(m_2)$ on the other.

Consequently the probability of generating the word forms in two languages can be decomposed into the product of generating the two language-particular forms for each meaning.

$$P(\lambda_1, \lambda_2|h) = \prod_{m \in M} P(\lambda_1(m), \lambda_2(m)|h)$$

4.2 Cognacy and independence

The next assumption holds that structural correlation between corresponding forms should be explained as resulting from cognacy.

Assumption 6 *Across different languages, forms corresponding to the same meaning are dependent only if the forms are cognate.*

If the words for a particular meaning do not derive from a common ancestral form, then they are uncorrelated. To return to our Polish and Russian examples, we can expect dependencies in structure between the cognate words **drzewo** and **дерево**. But we should expect no such correlation in the non-cognate pair **pomarańcza** and **апельсин** meaning ORANGE-nominative-singular.

Let us write M_i for the domain of the lexical function in language L_i. This is the set of meanings for which this language has defined a word form. The set of cognates is the domain of the lexical function of the parent language, M_0. We can decompose the evidential words into three sets: M_0 of cognates, $M_1 \setminus M_0$ of meanings only expressed in language L_1, and $M_2 \setminus M_0$ of meanings only expressed in language L_2. Words in the second and third categories are non-cognate, and so probabilistically independent of each other.

The conditional probability of the data can thus be expressed as follows.

$$P(\lambda_1, \lambda_2|h) = \prod_{m \in M_0} P(\lambda_1(m), \lambda_2(m)|h)$$

$$\prod_{m \in M_1 \setminus M_0} P(\lambda_1(m)|h) \prod_{m \in M_2 \setminus M_0} P(\lambda_2(m)|h)$$

4.3 Probability of a word

We now turn to the probability of generating a string in a language. The first assumption defines the distribution over word-length.

Assumption 7 *The probability of a word having a particular length is negative exponential in that length.*

The second assumption allows segment probability to depend only on the segment identity, and not on its neighbourhood.

Assumption 8 *Segment choice is context-independent.*

These two assumptions together imply that the probability of strings is determined by a fixed distribution over $\Sigma(L_i) \cup \{\#\}$, where $\#$ is an end-of-word marker. For the descendent languages, this distribution can be taken as the relative frequencies of the segments and end-of-word marker. Denote this distribution for language L_i by f_i.

The probability of generating a word in a language, given relative frequencies f_i, is the product of the relative frequencies for each lettern in the word, multiplied by the relative frequency of the end-of-word marker.

$$P(\lambda_i(m)|h) = f_i(\#) \prod_{a \in \lambda_i(m)} f_i(a)$$

Note that this expression only holds for words that are independent of all others, such as components of non-cognate pairs.

4.4 Probability of generating a cognate pair

The probability of generating a cognate pair of words is similar to the above, because descendent forms are deterministically derivable from the parent forms. If $(\lambda_1(m), \lambda_2(m))$ are a pair of cognates derived from an ancestral form $\lambda_0(m)$, then there is unit probability that the descendent forms are what they are given the parent: $P(\lambda_1(m), \lambda_2(m)|\lambda_0(m)) = 1$.

Since a cognate pair is derivable from a parent form, the probability of a cognate pair is the sum of the probabilities of all parent forms which will generate the two descendents. Write $W(m) = W(\lambda_1(m), \lambda_2(m))$ for the set of possible correspondence strings in the parent which project onto wordforms $\lambda_1(m)$ and $\lambda_2(m)$. Then the probability of the word pair is given by:

$$P(\lambda_1(m), \lambda_2(m)|h) = \sum_{s \in W(m)} P(\lambda_0(m) = s|h)$$

The summation poses a slight problem, however. How do we sum over all possible strings with given projections? Fortunately, we can decompose the summation. Start by recognising that

the parent language is also a language, and so the probability of forms in the language is determined by a distribution over segments — in this case correspondences — and the end-of-word marker. For consistency, we call this distribution f_0.

The only parent form which projects onto two empty strings is the empty string, consisting only of the end-of-word marker. For brevity, we will drop the lambdas, writing $P(x, y|h)$ for $P(\lambda_1(m) = x, \lambda_2(m) = y|h)$

$$P(0, 0|h) = f_0(\#)$$

We assume, without loss of generality, that the segmental inventory of the parent language consists of all degree-(u, v) correspondences between L_1 and L_2. Parent segments which are never used can be excluded by giving them zero relative frequency in f_0.

The function $Pre(s; u, v)$ returns the set of binary divisions (a, b) of the string s, such that the length of the first part a is at least u and at most v.

$$Pre(s; u, v) = \{(a, b)|ab = s, m \leq |a| \leq n\}$$

With this function, we can recursively define a function $W(s, t; u, v)$ on pairs of strings (s, t) which returns the set of all degree-(u, v) parent language strings which project onto s and t. For brevity, we will treat all u, v arguments as implicit.

$$W(0, 0) = \{0\}$$

By definition, the only parent language string which can map onto the empty string in both descendents is the empty string.

The recursive step breaks the strings s and t into all possible prefixes a and c respectively. The correspondence (a, c) is then preposed on all strings returned by W when it is applied to the remainders of s and t.

$$W(s, t) = \biguplus_{(a,b)\in Pre(s)} \biguplus_{(c,d)\in Pre(t)} (a, c)W(b, d)$$

Note that this is the set $W(m)$ we defined earlier.

$$W(m) = W(\lambda_1(m), \lambda_2(m); u, v)$$

The recursive definition of W in terms of disjoint unions and concatenation can be transformed into a recursive definition for the probability $P_0(s, t|h)$ of constructing a member of the set. Disjoint union is replaced by summation, concatenation by product. The probability of an individual correspondence (a, c) is its (unknown) relative frequency $f_0(a, c)$ in the parent language. Once again, we hide the implicit u, v parameters.

$$P_0(0, 0|h) = f_0(\#)$$

$$P_0(s, t|h) = \sum_{(a,b)\in Pre(s)} \sum_{(c,d)\in Pre(t)} f_0(a, c)P(b, d|h)$$

4.5 Probability of a form-pair

We now have the pieces to specify the probability of finding any particular form as the form-pair for the descendent languages. The probability of the pair in the case of cognacy is $P_0(\lambda_1(m), \lambda_2(m)|h)$. If the pair are not cognate, then they are independent, and their probability is $P_1(\lambda_1(m))P_2(\lambda_2(m)|h)$. If we write $c(m|h)$ for the likelihood that the pair is cognate, we can combine these two values to given a total probability of the two forms.

$$P_0(\lambda_1(m), \lambda_2(m)|h)c(m|h)$$

$$+P_1(\lambda_1(m))P_2(\lambda_2(m)|h)(1.0 - c(m|h))$$

Because the word-pairs are independent (assumption 4), the product of the above probability for each meaning m gives the probability of the data given the hypothesis.

4.6 Hypothesis

One burning question remains, however. What is the hypothesis? The simple answer is that it is exactly those free variables in the specification of the probability of the data

There were two groups of unknowns in the probability of the data. The first is the relative frequency f_0 assigned to correspondences in parent-language forms. The second is the likelihood of cognacy c, a vector of values between zero and one indexed by meanings.

A hypothesis is therefore any setting of values for the pair of vectors (f, c).

Note that while the degree variables u, v were not fixed in the above derivation, they will be held constant for any particular search, and thus do not define a dimension in the hypothesis space.

4.7 Search

In this section, we have derived $P(D|h)$, the likelihood of our data given a hypothesis.

For simplicity, we choose a flat prior over hypotheses, rendering the MAP Bayesian approach an instance of maximum likelihood determination. The value for the likelihood is differentiable in each of the parameters. Consequently, gradient descent can be used to find the hypothesis which maximises the probability of the data.

5 Results

In constructing the method, we made a number of assumptions about independence of forms. It is sensible that for testing, the method is applied to data that conforms reasonably well to these assumptions. The alternative is to apply it to data which contradicts its fundamental assumptions, consequently hampering its effectiveness.

5.1 The data

Polish and Russian were chosen to provide the data because they approximately obey assumption 6: words have dependent structures if and only if they are cognate. For our two languages, this means that borrowings from common sources are uncommon (numbering 45 in our data set), at least in comparison with the number of cognates (numbering 156).

The data was harvested from two online dictionaries (Wordgumbo, 2007a; Wordgumbo, 2007b), one English-Polish, the other English-Russian. Multiple translations were simplified, with the shortest translation retained. The English glosses were used as the meanings for the words. Where the gloss contained a capital letter, indicating a proper noun, this was eliminated from the data.

The data should also conform to assumption 4, that words for different meanings with a language are independent. So where two meanings in the data sets were realised with the same form, these meanings were deemed to be structurally dependent, and so only the first was retained in the wordlist.

The remaining data contains 407 aligned Polish-Russian word pairs.

Polish and Russian both use a great deal of derivational and inflectional morphology. The simple language model used here does not take this into account, so this will be a disturbing influence on the results.

5.2 Evaluation

The aligned wordlists were hand-tagged as cognate, common borrowing or non-cognate. A permissive rule of cognacy was used: if the roots of words in the two languages were cognate, they were cognate, even if represented with non-cognate derivational and/or inflectional morphology.

Figure 1 shows the evaluation of the program's performance on the data.

Borrowings as:	cognates	non-cognates
Found f	162	119
Missed m	41	37
Errant e	6	49
Accuracy $f/(f+e)$	96%	71%
Recall $f/(f+m)$	81%	76%

Figure 1: Evaluation of program performance on 407 meaning-matched pairs of Polish-Russian words. Common borrowings are scored as cognates in the first column, non-cognates in the second.

The scores show that the method works well in identifying cognates, particularly if common borrowings are accepted as cognates, or excluded manually. If common borrowings are scored as non-cognates, then the accuracy falls.

Of the correspondences found between Polish and Russian, 67 have a phonological basis. The remaining 27 result from mismatch morphology in cognates or differences in common borrowings.

6 Conclusion

This paper has presented a model of language which allows the calculation of the posterior probability of forms arising in the cases where

they are cognate, and where they are not. Bayes' theorem relates these probabilities to the posterior likelihood of particular correspondences and cognacy relationships. Gradient descent can be used to search this space for the best distribution over correspondences, and best cognacy evaluations for meaning-paired words. The application to data from Polish and Russian shows remarkable success identifying both cognates and non-cognates.

Future work will proceed by relaxing constraints on the parent language. The parent inventory will be widened to include multisegment correspondences. Multiple parent languages will be permitted, to the end of separating borrowings from cognates. Finally, richer models of language, incorporating syllable structure, will allow more information to identify cognates.

References

Gil Broza. 1998. Inter-language regularity: the transformation learning problem. Master's thesis, Institute of Computer Science, Hebrew University of Jerusalem, October.

Isidore Dyen, Joseph B. Kruskal, and Paul Black. 1992. An Indo-European classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82(5).

T. Mark Ellison and Simon Kirby. 2006. Measuring language divergence by intra-lexical comparison. In *ACL*, pages 273–280, Sydney.

T. Mark Ellison. 1992. *The Machine Learning of Phonological Structure*. Ph.D. thesis, University of Western Australia.

Donald G. Frantz. 1970. A PL/1 program to assist the comparative linguist. *Communications of the ACM*, 13(6):353–356.

Russell D. Gray and Quentin D. Atkinson. 2003. Language-tree divergence times support the anatolian theory of indo-european origin. *Nature*, 426:435–439.

Jacques B. M. Guy. 1984. An algorithm for identifying cognates between related languages. In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*. Available online as http://acl.ldc.upenn.edu/P/P84/P84-1091.pdf.

Jacques B. M. Guy. 1994. An algorithm for identifying cognates in bilingual wordlists and its applicability to machine translation. *Journal of Quantitative Linguistics*, 1(1):35–42.

Hans Heinrich Hock. 1991. *Principles of Historical Linguistics*. Mouton de Gruyter, Berlin.

Robert J. Jeffers and Ilse Lehiste. 1989. *Principles and Methods for Historical Linguistics*. MIT Press, Cambridge, MA.

Martin Kay. 1964. The logic of cognate recognition in historical linguistics. Technical Report RM-4224-PR, The RAND Corporation, Santa Monica, CA, September.

Brett Kessler. 2001. *The Significance of Word Lists*. CSLI Publications, Stanford, CA.

Simon Kirby. 2002. Natural language from artificial life. *Artificial Life*, 8(2):185–215.

Grzegorz Kondrak. 2002. *Algorithms for Language Reconstruction*. Ph.D. thesis, University of Toronto.

April McMahon and Robert McMahon. 2003. Finding families: quantitative methods in language classification. *Transactions of the Philological Society*, 101:7–55.

I. Dan Melamed. 2000. Models of translational equivalence among words. *Computational Linguistics*, 26(2):221–249.

Luay Nakleh, Tandy Warnow, Don Ringe, and Steven N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an ie dataset. *Transactions of the Philological Society*, 103(2):171–192.

D. Ringe, Tandy Warnow, and A. Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society*, 100(1):59–129.

Wordgumbo. 2007a. /ie/sla/pol/erengpol.htm. Website http://www.wordgumbo.com/.

Wordgumbo. 2007b. /ie/sla/rus/erengrus.htm. Website http://www.wordgumbo.com/.