CERIAS Tech Report 2003-25

PARONOMASIC PUNS: TARGET RECOVERABILITY TOWARDS AUTOMATIC GENERATION

by Christian F. Hempelmann

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907

PARONOMASIC PUNS:

TARGET RECOVERABILITY

TOWARDS AUTOMATIC GENERATION

A Thesis

Submitted to the Faculty

of

Purdue University

by

Christian F. Hempelmann

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2003

for Miriam

ACKNOWLEDGEMENTS

I wish to acknowledge the many people who were of assistance to me for the present project. There are those who provided many useful comments at the presentations of parts of this research in Bertinoro, Hannover, Nijmegen, Trento, and Enschede. My fellow graduate students in the Interdepartmental Program in Linguistics and the Program in English Language and Linguistics at Purdue University also listened patiently as I was developing and explaining my ideas at our weekly *Stammtisch*. Special thanks are due to Donca Steriade and Heidi Fleischhacker, who made their research available to me, and in particular to Wlodzimierz Sobkowiak, who provided me with his tagged database of puns which lies at the core of my empirical testing and without which this project would not have been possible. Krista Bennett not only translated some of my results into actual code, but also answered my many inept questions about it.

The project also benefited greatly from comments and suggestions of my committee members, Salvatore Attardo, who has accompanied me through much of my academic career, Shaun Hughes, who has a marvelous eye for details, Mary Niepokuj, from whose phonology class the original idea of this paper developed and with whose help it grew into the present study, Sergei Nirenburg, and, of course, the chair of my committee, mentor and *spiritus rector*, Victor Raskin, who supported me in too many ways, linguistic and other, to even think of mentioning them. Part of this research was supported by IIE grant no. F103000-15013790 and by the Eli Lilly Foundation establishment grant to the

Purdue Center for Education and Research in Information Assurance and Security (CERIAS) at Purdue University.

TABLE OF CONTENTS

Page

ABSTRACT viii

1. INTRODUCTION 1
    1.1 Outline 1
    1.2 Main Hypotheses 4
    1.3 General Linguistic Parameters of Punning 9
    1.4 Conventionality of Punning Similarity 17

2. LINGUISTIC HUMOR THEORY 20
    2.1 Semantic Mechanisms of Humor 20
    2.2 The Illogical Logic of Puns: Cratylism 23
    2.3 Pseudopunning Wordplay 31

3. PHONETIC DISTANCE IN TERMS OF DISTINCTIVE FEATURES 34
    3.1 Vitz and Winkler 1973 35
    3.2 Zwicky 1976 37
    3.3 Zwicky and Zwicky 1986 39
    3.4 Sobkowiak 1991 43
    3.5 Summary 44

4. TARGET-PUN DISSIMILARITY AS OUTPUT-OUTPUT CONSTRAINT VIOLATIONS 46
    4.1 Optimality Theory 46
    4.2 Correspondence (McCarthy and Prince 1995) 51
    4.3 Output-output Correspondence (Benua 1997) 52

Page
4.4 Perceived Similarity and Correspondence Environments (Steriade 2001)..... 56
4.4.1 Theoretical Note..... 56
4.4.2 The P-map..... 58
4.4.3 Correspondence Domains: Rhymes and Puns 59
4.4.4 The C-Constraint: Multiple Survivors 61
4.5 Onset Correspondence in English Imperfect Puns (Fleischhacker 2002)..... 64
4.5.1 Perceptual Similarity vs. Simplification..... 64
4.5.2 Again: Punning Similarity is more than Acoustic Similarity..... 66
4.5.3 More Frequent Violations = More Similar Puns?..... 67
4.5.4 More Similar Puns = Better Puns?..... 68
4.5.5 Onset Difference in Heterophonic Puns..... 69
4.5.6 Not CONT, but MAX (and DEP) 71
4.6 Similarity of Personal Names (Lutz and Greene 2001)..... 73

5. PUNNING CORRESPONDENCE CONSTRAINTS 81
5.1 Overview of Previous Results 81
5.2 Asymmetrical OO-Correspondence in Punning..... 84
5.3 Constraints on Punning OO-Correspondence 86
5.3.1 Phoneme Classes..... 88
5.3.2 Acoustic Feature Tree 89
5.3.3 Vowel Space 93
5.3.4 Environment of Constraint Violations 96
5.4 Suprasegmental Constraints 97
5.5 Cumulative Cost of Constraint Violation 98
5.6 Evaluation of Punning Output-Output Correspondence..... 100
5.7 Implemented Edit Distance Computation..... 101
5.8 Overcoming the Infinity Problem of OT 103

6. RECOVERABILITY BEYOND PHONOLOGY 105
6.1 Introduction: Interaction of Phonological, Syntactic, and Semantic Factors.. 105
6.2 Quasi-punning non-phonological Recoverability..... 108

Page
6.3 Intermediate Level: Morphology 109
6.4 Superlexemic, Multiple-Word Domain Punning 111
6.5 The Context as Torso 112
6.6 Idiomatic Expressions: Semantic-Syntactic Interface 115
6.7 Word Classes and Syntactic Functions: Lexico-Semantic Interface 121
6.8 Recoverability through Ungrammaticality 127
6.9 Syntactic Ambiguity 130
6.10 Definition Puns: Less Semantics, Less Joke 132

7. SUMMARY AND OUTLOOK 134

BIBLIOGRAPHY 136

APPENDICES 146
Appendix A: CLAWS Tag Set 146
Appendix B: Syntactic Function Tag Set 149
Appendix C: Consonant Distributions and Frequency 150
Appendix D: Non-Corresponding Segments 152
Appendix E: Main Corpus of Target-Pun Pairs 166
Appendix F: Intrusion Frequency Table 186
Appendix G: Cost Table 187

VITA 188

ABSTRACT

Hempelmann, Christian F. Ph. D., Purdue University, August, 2003. Paronomasic Puns: Target Recoverability towards Automatic Generation. Major Professor: Victor Raskin.

The aim of this dissertation is to create a theory to model the factors, prominently, but not exclusively the phonological similarity, important in imperfect punning and to outline the implementation of this measure for the evaluation of possible imperfect puns given an input word and a set of possible target words. Imperfect, heterophonic, or paronomasic, puns differ from perfect, homophonic puns in that the target is different in sound from the pun. While homophonic puns are interesting for the linguist primarily with respect to their semantics, heterophonic puns present a research issue also to the phonologist, because they use one of two similar sound sequences to stand for both meanings associated with them, for example, *bang* to denote a noise as well as a financial institution. The specific question here is, how much contrast is possible between the pun and its target to make the latter recoverable, in terms of the semantics, phonology, and syntax of the pun-target pair and its context. The theoretical framework for the phonological part of this project is inspired by a recent version of Optimality Theory (OT), adopted in phonology, because it is able to describe the occurrence of related forms through a selection process from among possible candidate forms more appropriately than derivational approaches can by way of rules operating on one input form and

yielding one output form. Taking more parameters—both phonological and syntactic—into account than previous studies, this project is intended to describe the linguistics of the imperfect pun in terms of a set of hierarchies of constraints weighing the differences found between the puns and targets of a sample corpus. Based on this measure, I will outline a computational implementation of the results that can evaluate an input word with respect to a set of existing English words from a machine-readable dictionary. The basic idea is to assign values to constraint violations and combinations thereof and then adding up the "penalty" for each violation in which a possible target does not conform to the pun.

1. INTRODUCTION

1.1 Outline

The intended audience of the present study are primarily humor scholars, for whom the basic linguistic ideas involved will be outlined in sufficient detail, but also linguists—prominently, phonologists—who should be interested in the insights gained from applying their linguistic theories the field of humor and for whom so much detail will be provided in some sections that the humor scholars may not be able to benefit from all of them, and I apologize to my colleagues in the field of humor research. This is particularly the case in chapters 3 through 5. The humor scholars, in turn, may want to skip section 2.1 as it introduces basic semantic humor theory, which will be relevant for the linguist to follow the discussion.

Humor in general and puns in particular, the central focus of the present study, provide a field of application for all subdisciplines of linguistics, and chapter 6 will focus on linguistic contributions to the study of the pun beyond phonology. But semantics, in particular script-based semantics, has contributed most prominently to the analysis of humor. Accordingly, after the phenomenon of the pun as the purview of the paper has been outlined in chapter 1, this contribution will be presented in detail in chapter 2. On this foundation, I will focus on the question what makes a good heterophonic pun in the remaining chapters of the present research.

Chapter 3 surveys previous work on sound similarity that approached the issue at the level of distinctive features in which two segments differ. While there are some distributional insights to be gained from such approaches, it will become clear in my discussion that the isolated distinctive feature is neither the degree of abstraction, nor the amount of environment—namely none—with the help of which the relation between a pun and its target word can be described. But important empirical results from Sobkowiak's study (1991) will be used to inform the present research inspired by the more recent and theoretically much better grounded work in optimality theory (OT) surveyed in the following chapter.

In chapter 4 the theoretical framework for this central section is a metaphorical model-extension of a recent version of OT, which will be briefly summarized in its canonical form, before the following sections will focus on Correspondence in OT and its theoretical implications as well as several applications. Recently, OT has been very successfully applied to the relation between related output forms, for example, in reduplication and semi-rhymes, as well as in an initial study on the onsets of imperfect puns.

Chapter 5 is the core of this research, where the previous discussion and results are brought together, adapted and united with the new insights gained in the present study. Taking more parameters—phonological, syntactic, and semantic—into account than previous studies, and accounting for the environment of the differences as well as the asymmetricality of the target-pun relation, my project is intended to arrive at a hierarchy of constraints weighing the differences found between 1,182 puns and targets of a sample corpus (selected from 4,000 transcribed puns in Sobkowiak (1991) and based on Crosbie

(1977)). The format of this central section is the detailed description of the output-output constraints violated in punning and the cost implications of these violations. Based on this measure, the final section of chapter 5 outlines a computational implementation of the results for an application of this research that can evaluate an input word with respect to a set of existing English words, for example the CELEX corpus, or CMU Dict. The basic idea that will be developed in this section is to assign values to constraint violations and combinations thereof and then to add up the "penalty" for each violation in which a possible target does not conform to the pun. The lower the penalty, the better the pun. If the penalty is too high, the target is out.

Chapter 6 shows how these phonological constraints of imperfect punning interact with supporting phenomena analyzed at the morphological, syntactic, and semantic levels. Syntactic parameters that need to be taken into account are the word class identity of pun and target, as well as the question across what kind of word boundaries and in what syntactic phrase contexts such contrast between these forms can be established, as, for example, in the following attested puns: *Adelaide* targeting "I'd have laid"[1] and *Adelia* targeting "I'll deal you." In particular, formulaic patterns like idioms and proverbs have been found to play a prominent role at this level. Discussing the importance of the interaction of such issues casts a new light on the relative importance of phonological similarity in punning as a type of ideal, symbolic, and not primarily acoustic similarity which functions as the logical mechanism of the humor in jokes, and at the same time concrete, acoustic similarity that facilitates the recovery of a target given the word that

[1] I am grateful to Shaun Hughes who made me aware of Sobkowiak's mistake in reading the target of this puns as "I've laid."

puns on it and the torso of text surrounding it. As such it describes new insights into the pun as a linguistic problem in general.

The final summary in chapter 7 provides an overview of the results achieved in this research, and, on this basis, approaches the theoretical issues of drawing clear boundaries in the purview of research, and the implications of the present research in humor theory for linguistics in general and phonology in particular.

1.2 Main Hypotheses

The central question that inspired this research is whether the phonological difference between a pun and its target can be described and explained more adequately within a framework inspired by Optimality Theory (OT) than it was possible with previous theories, most notably feature-based approaches. While this will shed light on the nature of puns as the particular object of study, such an explanation will also facilitate a better understanding of the phonology of English in general, as well as the usefulness of the theoretical approach and its method.

Trivially underlying this question is the observation that a pun and the word it is punning on must be similar in sound. Deriving from this observation and research in humor theory, the present study will be based on two distinct hypotheses related to the sound similarity in imperfect puns and just as valid for the sound identity in perfect puns:

(1) A target and its pun cannot be arbitrarily different in sound because their similarity has a function in
   a. the phonological support for recovering the target from the pun, and
   b. creating the humorous semantic overlap between the two scripts, one including the pun, the other including, and often triggered by, the target.

In the following example, with respect to hypothesis (1a) the target "pearly" is sufficiently similar in sound to be recoverably from the pun "curly" to evoke the latter, and the idiomatic force of the collocation "Pearly Gates" aids this recovery substantially.

(2) *Labia majora*: the curly gates.
curly -> pearly (60)[2]

Note the nomenclature used in the present study. The term *pun* will be reserved for the ambiguous word in the text—"curly" in example (2)—of what will be called the *punning joke*. The second word that sounds similar to the pun and is not actually present in the text, but paradigmatically available—is called the *target*—"pearly" in the example.

With respect to hypothesis (1b), the two main semantic requirements for a text to be a joke are, in short, a semantic opposition, here the sexual vs. non-sexual as non-religious vs. religious, in general a low vs. high opposition, and compatibility of part of the text with both these meanings. This portion of the text is the pun, compatible with one meaning, and the paradigmatically available target, compatible with the other. But there is not only the overlap in sound between the pun compatible sexual meaning ("curly" pubic hair) and the non-sexual religious meaning triggered by the target ("pearly" gates), but also a semantic one, both "gates" being entrances to pleasurable, blissful, places. In this respect, despite being somewhat crudely anatomically graphical, example (2) is actually an exceptionally good punning joke. As will be discussed in more detail in section 2.3, many puns create the overlap only through the sound similarity, which

[22] I adopted the following notational convention for pun examples in these early chapters where examples aren't taken exclusively form Sobkowiak (1991): After the full text, first the pun and, after an arrow, the target are given in italics as they are spelled in English; after a comma, the standard IPA is supplied in the usual square parentheses for both, pun and target, if the phonology is part of the point of the particular example; in later examples the transcription is provided according to Sobkowiak's convention after it will have been introduced; if the source of the pun text is given as a number only, it refers to pages in Crosbie (1977), the source of the corpus used here.

renders them mere wordplay, but not humor. And I claim that this lack of semantic overlap is the reason puns are often perceived as feeble and unsophisticated humor.

In sum, if a pun in a text is too different in sound from the target to fulfill function (1a), the punning joke fails completely, no humor is created, the text is not a joke, and, if the attempt to joke has been detected, the teller will probably be prompted to supply additional explanations to make the target recoverable, although this is a dangerously face-threatening act. But if the pun and target are sufficiently similar in sound for the latter to be recovered, the text may be perceived as a joke, but the pun-target pair may still be so dissimilar in sound that function (1b) is barely fulfilled and it will be perceived as a bad joke. But, again, this second function is interacting in crucial ways with the semantics of humor.

As should have become clear, the recovery function (1a) of the sound similarity will be the central interest of this dissertation. The specific kind of punning logical mechanism (1b), on the other hand, is discussed in detail in section 2.2. This specific form in which the major requirement for a text to be funny is manifested in punning is a major factor interacting with the other ones I will introduce in the present study. At one level, the semantic relation between a pun and its target is straightforward: two meanings of words are similar, because their signifiers sound similar. But while I will focus on the question in what way they *sound similar*, I will not attempt the far more interesting task of explaining how the meanings *are* alike, "overlapping" in terms of the semantic theory this approach is based on, beyond summarizing and relating what has been attained so far, and what may well be all we can attain at this stage of research in semantics. The assumption is that only if puns conform to the phonological constraints that will be

discussed here, can they fulfill their semantic function, the latter being, so to speak, a constraint working largely against—and not alongside—the ones within the scope of the theories employed here. While this sounds trivial, it has the important implication that the humorous logic is indispensably facilitated by the phonological similarity. In other words, puns and targets must be as similar as possible to fulfill the functions described in (1), while they must be different enough to represent two different lexical items denoting two different meanings, which can then be presented as opposite in the text of the joke. But they can never be perceived as not being similar at all to still form a pun-target pair.

A related, but different issue is the semantic support for the recoverability of the target. This rarely operates along the lines of the semantics of humor in particular, but through the general principles of semantic coherence and priming. This point will also be discussed in more detail in a chapter 6 in connection with the non-phonological factors of heterophonic punning.

The main problem that arises from this broader look at the issue of punning is the delimitation of the purview of the research in order to not have to do a full analysis of *all* aspects of heterophonic punning, or—even worse—any text in the English language. A joke, and in particular a verbal joke, of which the pun is the prototype, is a text, an artifact of the most complex symbolic system we know, namely, human language. In the processing of a text, we employ all techniques available to us, and these techniques are analyzed in linguistics in different subdisciplines, most centrally semantics including pragmatics, but also phonology, morphology and syntax. Therefore, an exhaustive analysis of a text will use all available theories and methods from all these fields to capture all factors, as well as the complex interaction between them, relevant to the texts

description. But such an approach is, of course, neither science nor attainable. A level of abstraction must be found at which we can make meaningful statements based on the theory we chose, so that we can ignore less relevant issues and ensure that our model of the phenomenon under scrutiny does not aim to recreate the phenomenon itself and thus become just as unwieldy and consequently fail to provide insights.

I realize that I open myself up to this very criticism, when I extend the discussion to the syntactic and semantic factors of heterophonic punning. But I claim that failing to outline them to illustrate the importance of these aspects for the phonology of imperfect puns would be equivalent to discussing the history of Bauhaus architecture based exclusively on the basis of the types of concrete used in erecting the actual buildings and a general notion of angles. In short, while this paper is about sounds as the stuff from which language is built, it needs to relate its findings to the purpose for which it is used, namely to create meaning—understandable to humans practically and functioning aesthetically, as English language and as the art of humor.

The rationale underlying my approach is then still first and foremostly a phonological one: Heterophonic puns can reveal much about phonotactics, because they have to obey not just them, but also the need for one segment to sound like one word (or lexeme) and at the same time close enough to another, while homophonic pun segments sound exactly like two lexemes. The question to be answered is then, in short, what *sounding similar* means and *how* the similarity is realized in pun-target pairs. The answer will lie in a descriptive method inspired by OT and an explanation sketched in relation to perceptible similarity and additional crucial factors lying outside the scope of purely phonological approaches.

1.3 General Linguistic Parameters of Punning

After the theoretical issues involved in this research have been briefly introduced, the ground is prepared for a more detailed description of the phenomenon at hand. In view of the rich taxonomies developed very early on in classical poetics and rhetoric, it seems helpful to have a brief look at medieval rhetoric as it can sufficiently contribute to the clarification of terminological issues in puns. The *Rhetorica ad Herennium* (Ad Her.), long attributed to Cicero, and the *Poetria Nova* (PN), attributed to Geoffrey of Vinsauf and based on the doctrine of Ad Her. (cf. Gallo 1971), include the following selection of relevant figures of word-play:

1. *traductio*:[3] same sign, used twice (*Ad Her.* IV. xiv. 20-21, *PN* 1104-8)
2. *adnominatio*: slightly different sign, used twice (*Ad Her.* IV. xx. 29, *PN* 1140-44)
3. *significatio*: same sign, used once (*Ad Her.* IV. liii. 67, *PN* 1550-53)

Medieval rhetoricians—as well as modern-day linguists as we will see in this section—put emphasis on the distinction between homonymous (mostly homophonic and

[3] A special form of *traductio* is the *rime riche*. Ideally, a rime riche uses the same sign not only twice, but also in rhyming, usually endrhyming, position. The signs should be identical in sound, but in less ideal form also only similar. Rime riche, like the other figures of speech, need not necessarily have a punning, that is, humorous, intent, but can be a purely rhetorical ornament or a style-marker. The purest form of the rich rhyme is then the connection of two phonologically identical forms that stem from different words and have different meaning.
So that I have my lady in myne armes.
For though so be that Mars is god of armes, (KnT I 2247-48)
[armes = arms (limbs + weapons)]

possibly homographic[4]) puns and heteronymous, heterophonic, paronomasic, or "imperfect puns." The former feature the same sign once, or twice, in both cases with two unrelated[5] meanings, the latter feature two slightly different signs, or one sign that triggers a slightly different one, both with different meanings. The following table illustrates the difference:

Table (1): Types of Punning

A. heteronymy		[horological instrument vs. male genitalia]
1. single sign	(pure)	"your clock is very big"
2. double sign	adnominatio	"his clock was bigger than his cock"
B. homonymy		[domesticated male bird vs. male genitalia]
1. single sign	significatio / syllepsis	"the farmer has a big cock"
2. double sign	traductio / antanaclasis	"the cock has a big cock"

In this table, types of pun are listed according to two main criteria and two intersecting additional criteria: Heteronymical puns under A involve a pun and target that are not identical but similar in sound, while homonymical puns are a special case of this in that the similarity is perfect, that is, pun and target are identical in sound. Under each of these categories a distinction is made according to the presence or absence of the

[4] Homophony and homography are subcategories of homonymy: Two homonyms share spelling and pronunciation, while homophones share only pronunciation and homographs only spelling. Before Caxton's introduction of the printing press we find no large-scale systematization of spellings in late Middle English, thus orthography lent itself to the molding hand of the inventive scribe and homophony was more important as a criterion for potential puns than today. When a certain spelling indicated phonetic differences it could even be changed to some degree to function properly, for example, for end rhymes: *reaume, reawme, reame, realme, rewne* ("realm"). The distinction between *adnominatio* and *traductio* is so much at the mercy of the Middle English author (or scribe) that he or she can turn an *adnominatio* into a *traductio* at his—or, rarely, her—will. The same goes for the distinction between pure heteronymous puns and *significatio*. This is not to say, of course, that pronunciation—and accordingly spelling—was so free that "anything went:" Chaucer's beautiful *adnominatio* with "frankes" and "flankes" in ShipT VII 201-02, for example, could not have become a *traductio* (except for an audience consisting of native speakers of a language in which the distinction between the liquids [r] and [l] is not phonemic, for example, Japanese.

[5] When the two meanings are different, but related, one speaks of polsysemy, e.g., in the metaphorical extension of the meaning of "hot" from referring to a temperature to denoting an attractive human. For the present discussion the distinction between homophony/heterophony and polysemy is not assumed to be a significant parameter, as it seems to have little or no effect on the humor of the pun if the two meanings are etymologically related or not. I am grateful to a colleague at the University of Twente who made me aware of the possible significance of this parameter, but will leave the pursuit of this issue to others.

second sign, the target, in the actual joke text (2) or only paradigmatically as triggered by their corresponding puns (1). In type A1 a single sign, the pun, is used to evoke a second similar sounding sign which is thus paradigmatically present. A less frequent variant of this typical kind of paronomasia is illustrated in row A2 providing an example of *adnominatio* in which both the pun and the target are present in the text. This form of punning is not virtually never used in contemporary humor. Under B1 we find the corresponding variant of punning where only one sign is present in the joke text but it evokes the second identical sign and both their meanings, and under B2 the variant that shows both identical signs present in the text.

It should be clear why *significatio* (B1) is considered to be the prototypical, "pure" type of pun: Using only one sign, two opposed meanings are denoted that share this sign completely. But this paper is restricted to the discussion to puns that are characterized by heteronymy in absence—or rather virtual presence as the target—of the second sign. The example in the table (A1) features the pun-target pair *clock-cock* with only the pun actually present.[6] This is the most frequent type of punning occurring in English today.

Zwicky and Zwicky summarize as follows: "In a perfect pun, a single phonological entity is to be understood as representing two distinct lexical items. In an imperfect pun, one phonological entity (the p u n) stands for a phonologically distinct (but similar) entity (the t a r g e t), the two representing two distinct lexical items as before." (1986: 494). So in fact, the imperfect pun *means* both, the pun and the target. And while in a

[6] Although this example was concocted for illustrative purposes, consciously aiming for a sexual script triggered by the target because of the universality and prominence of these scripts, this actual pun-target pair is too good (as a pun, if not as a joke) to be passed up by Crosbie (1977: 49), although not in its sexual sense, but figuratively for dumbness: "For years he stayed at home and collected clocks. Then, one day when he was sixty-nine, he finally admitted his real problem was that he was a clock sucker."

perfect, homophonic, pun phonology has no contribution to the understanding of this overlap of two meanings, in imperfect, heterophonic, puns we can describe how much distinctness is allowed and how much similarity is required. In sum, the field of application of this dissertation are English heterophonic, imperfect, or paronomasic puns. Before I will go into a finer taxonomy of these below, let me briefly distinguish them in more detail from perfect puns.

In homophonic puns there is (almost) zero phonemic distance. The following is one of the rarer examples of word boundaries being involved, a phenomenon I will return to in chapter 6. The punning segment consists of the sequence [əwɛfr], the pun is *a wafer* and the target *away for.* Note that the distinction between pun and target is not clear in homophonic puns, as the signifier is fully compatible with both signifieds in the given context, but a web search on this pun yielded an almost unified result for *a wafer* as the actual pun in the text.[7] The distinction is always clear for heterophonic puns.

(3) Why did the cookie cry?
Because her mother has been a wafer [əwɛfr] so long.
a wafer ->away for
(from Pepicello and Green 1983: 59, quoted in Attardo 1994: 128)

In heterophonic puns not all phonotactically possible distance is encountered. For example, the English word *hit* is obviously not a possible pun on the target *proprioceptively*, to pick two radically different, but possible sound sequences of English. In a less far-fetched and rare instance found in *The Onion*, a U.S. American satirical magazine, the fine line that distinguishes sufficiently similar from insufficiently similar

[7] This result was to be expected, because the script for BAKING GOODS needs to be firmly established and confirmed through *wafer* to form the background for the switch to ABANDONED CHILD which demands playful resolution in the answer to the question why the cookie cried (see chapter 2).

pun-target pairs is topicalized: The article makes fun of the sophomoric humor of parochial adolescents who refer to the towns neighboring their hometown through imperfect puns. For example, fictional Paul Sadecki from real Vandalia, Illinois, refers to St. Elmo as "St. Smellmo" and Litchfield as "Bitchfield." But when it comes to another nearby town, Van Burensburg, Sadecki runs into problems:

> "Van Burensburg seems like it'd be easy, but it's not," Sadecki said. "I used to call it Van Turdburglar, 'cause they got a lot of homos there, too, but *that kind of felt like a stretch*." ("Man has Derogatory Nicknames..."; my emphasis)

This example is somewhat atypical, because the pun is a name, a class of words for which special phonotactic rules hold (cf. section 4.6 below). But the main point here is that the need to refer to two different concepts, one denoted by the existing English word "turdburglar" (derogatory for homosexual) and the other the name of a town named after a President of the United States, which are in a humorous constellation for Sadecki—presumably high vs. low register—, force him override the sound similarity requirement to a degree that the joke may still be perceived as such, because the target is clear enough in the given context, but only as a bad joke because that target does not sound similar enough to the pun, the relation between them "kind of felt like a stretch."

These observations imply a cut-off point, above which we have possible puns and below which we have two segments not perceived as (sufficiently) related in sound to make the target recoverable at all, an issue I'll return to in section 4.4.4. But even above this cut-off point, there is better puns and worse puns, even in phonological terms, that is, puns that sound more similar to their target and those that do less so. According to my

(admittedly non-native) intuition, the following pun doesn't work too well (the next three examples are from Monnot 1981: 91ff):

(4) Public transportation: It's a dime good deal.
dime -> *damn*, [daym] -> [dæm]

I think the following two are better:

(5) Schmidt Beer: Rhapsody in brew.
brew -> *blue*, [bru] -> [blu]

(6) Thirst things thirst: Miller.
thirst -> *first*, [θrst] -> [frst]

Of course, the central assumption for this paper is that this intuition is motivated largely by the smaller distance, or reversely, a greater similarity of sound in the latter examples. While in (4) we have a vowel targeting a diphthong, in (5) we have one lateral targeting another, and in (6) one fricative targeting another, their difference in the given environment supposedly being smaller than that between the vowels. This observation is curious as, generally, the consonantal skeleton is more important for the identity of a word than its meat of vowels, es cen be seen frem thes semeler exemple en wretten Englesh.[8] I will return to this issue in section 5.3.

But Sobkowiak observes correctly that a main issue is the "the non-categoricality of most punning phenomena, where the traditional all-or-nothing, or yes-or-no approach is bound to fail" (1991: 3). And OT is the prime candidate for a such a gradational, shades-of-grey, analogue approach. Application of an extension of OT aimed at describing how much difference is possible in semi-rhymes (Steriade and Zhang 2001) and how much variability can be tolerated in proper names (Lutz and Greene 2001) can be extended to

[8] I must give credit to Mayerthaler for the original play with letters along the lines used by me here (1982: 226n).

answer a closely related question that is posed by heterophonic, or paronomasic, puns, and this has been partially attempted in Fleischhacker 2002. In sum, the two central question are, where the threshold lies between puns that do not identify the word they pun on, and puns that do (Attardo 1994: 121) and, in addition, to sort the latter according to higher sound similarity. In other words, how much distance is permissible for paronomasic pun-target pairs to still fulfill their task of representing two words and two meanings in one signifier, and which possible pun-target pairs do so better than others?

Earlier approaches have measured the distance in sound of two words in terms of the number of distinctive features in which two forms differ (e.g., Vitz and Winkler 1973, Sobkowiak 1991, Frisch 1996). Comrie illustrates this type of construal:

> "[T]he phonetic distance between two segments may be defined as the number of features (within some framework of overall features) by which they differ; if features are arranged hierarchically [...], then hierarchical position will also be relevant, with alternation of a feature higher in the hierarchy involving greater phonetic distance." (1979: 51-52)

And for the discussion of sound distance, pun-target pairs are as fertile a field of application as are related issues which have been discussed on the basis of distinctive features, e.g., semi-rhymes (Zwicky 1976) speech errors (Fromkin (ed.) 1973, 1980; Frisch 1996), or approximations in aphasia (Joanette, Keller, and Lecours 1980). To the degree that these issues can inform my approach, the literature on them will be surveyed in chapter 3 below.

The main difficulty for these feature-based approaches is pointed out by Steriade:

> When we assess the relative similarity of two pairs x-y and w-z, the simplest case is that in which the pair x-y shares a number of properties, and the pair w-z shares those same properties plus others. The less similar pair, x-y, shares a proper subset of the features common to w-z. If however the shared properties of the two pairs *do not stand in a subset*

> *relation*, the evaluation of relative similarity poses an obvious difficulty: is the similarity comparison meaningful in this case and, if so, what do speakers compare to find the more similar pair? (2001: 7; my emphasis)

Distinctive features are a problematic tool for the kind of comparison required to assess pun-target similarity. They are largely based on articulatory criteria and do not form a hierarchy of quantifiable relations, witnessed by the fact that, for example, the fricatives [f], [s], and [ʃ] are hardly distinguishable from each other, because they differ only in the place of articulation in the SPE system (Chomsky and Halle 1968), while the same holds for [f] and [x], which are clearly perceived as more different, regardless of their environment, implying a higher cost when they are the only difference between a target and its pun. An improvement over these approaches surveyed in chapter 3 lies in the use of features as arranged in the feature geometry model. But more importantly, the relative distance values assigned to featural difference must be based on their acoustic effect, which is highly dependent on the environment in which the difference is located, an issue related to grouping sounds into natural classes according to identical behavior in identical environments. I do not pursue this idea any further here in this introduction, but will return to this issue in the context of Steriade's *P-map* hypothesis discussed below in section 4.3 and the specification of my own parameters for the constraints on pun-target similarity relation in chapter 5.

Another issue of early feature-based approaches to imperfect puns is the alignment of the domains that are compared. To optimize correspondence between a pun and its target it is often necessary to start the comparison not at either edge of the words, or to skip word-internal segments in the comparison. This is illustrated by Vitz and Winkler in the following example (1973: 375):

(7) sɪt* *sit*
*ɪts *its/it's*

I will return to their approach to alignment in more detail in section 3.1. A full implementation of the results of the present study will require the automatization of alignment with the help of Markov-chains.

On the basis of a more detailed discussion of these issues, the present study will turn to the contribution that an OT-inspired theory can provide to a model for the central issue of a punning similarity model in English. The very struggle on which OT focuses, namely between opposing forces cast in constraints to evaluate several outputs with respect to one input finds a special correspondent in heterophonic punning. Namely, that between sufficient FAITHFULNESS[9] of the pun to the input target in order to render it recoverable and MARKEDNESS to indicate meaning differences required by the humor the text has to carry, when transferred into output-output (OO) constraints holding between two surface forms. As noted above, puns depend on confusability, i.e., high similarity. The motivation for deviance, or MARKEDNESS, is the pressure to represent two underlying and recognizable forms at once, as well as conforming to phonological well-formedness.

1.4 Conventionality of Punning Similarity

It has to be heeded that there is no a priori identity or even similarity in the processes, or constraint hierarchies, that underlie puns and other "metaphonological" phenomena. Zwicky observes this for the phenomena in his purview: "the principles of English rock

[9] Small caps are used both for OT constraints as in this instance (see section 4.1) and semantic scripts (see section 2.1). Since they will not occur in the same points of the discussion, no ambiguity will arise from this.

rhyme do not follow in any obvious ways from other relationships between segments ... we see artistic conventions using some rather abstract aspects of phonological systems" (1976: 694; cf. 3.2 below). That is, there are types of sound similarity that cannot be recorded with a microphone and reduced to an algorithm based on acoustic principles alone. As with all matters that involve language, which implies human users, the situation is more complex than purely physical or biological models can capture. Language is always also a cultural artifact. That we can expect less aesthetically informed arbitrariness in the similarity relation between a pun and a target lies in its necessary recovery function which must be assumed to allow for less variability.

Another example of this conventionality of similarity can be seen in Irish syllabic poetry which shows an obviously culturally specific concept of rhyme (cf. Knott 1974). Here, a rhyme between words is considered perfect, *comhardadh slán*, when "the stressed vowels are identical, and all the consonants subsequent to the first stressed vowel are of the same class and quality" (Knott 1974: 4). These classes appear quite heterogeneous to members of cultures not versed in Middle Irish poetry. The sounds represented by the letters *bh, gh, dh, l, mh, n, r*, for example, are considered to rhyme, as well as *ll, mm, ng, nn, rr*. Accordingly, *cluineam* and *fuigheall* are in a relation of perfect rhyme, a judgment clearly different from contemporary concepts of rhyming in English, where much less variation is allowed (cf. section 4.4.3). And since both corresponding forms in rhyme are part of the surface text—in contrast to imperfect punning where the target is only paradigmatically available and has to be recoverable from the pun and context—we can assume that rhyming indeed allows for less sound similarity between forms and the notion of similarity to be subject to culturally specific aesthetic influences.

Also, while in puns there is obviously a different additional motivation beyond the sound similarity in semi-rhymes, namely the semantic one sketched in section 1.2 and discussed in more detail in section 2.3, for both we have the general structure of two corresponding forms, one, the pun, being dependent on the other, the target. This goal is the “perfect rhyme” for semi-rhymes, and the homophonic target for paronomasic puns. For both, the general question is the same, namely, how much deviation from the ideal goal they tolerate, even if deviation must clearly be considered in terms of different parameters for each domain.

2. LINGUISTIC HUMOR THEORY

Before I continue the phonological discussion by surveying feature-based approaches to sound similarity in chapter 3, it is necessary to discuss the semantics of humor in general and the cratylistic logic of the metaphonology of puns in particular in sufficient detail in the present chapter. This will serve to ground the discussion in the context of linguistic humor theory, in particular the General Theory of Verbal Humor (GTVH; Attardo and Raskin 1991). Based on this, central parameters of my analysis can be derived from the groundwork laid in that theory, while my results in turn can inform the reshaping and enhancement of this theory towards greater descriptive and explanatory adequacy. Centrally, the importance of semantic analysis for a meaningful analysis of humor will become clear. In addition, the entangled interaction of linguistic factors in the verbal humor of puns, repeatedly pointed out above, centrally in the main hypothesis (1), will become more formalized and applicable when cast into the terms of a model that is based on tried and tested linguistic humor theory.

2.1 Semantic Mechanisms of Humor

The present approach is eclectic in the sense that it selects theories on the basis of the usability of the methodologies within those theories for the purpose of answering the main issue. While the phonological part will be formulated following a optimality-theoretic framework, the humor-theoretic part of it is based largely on the linguistic

theory of humor as developed in Raskin's Semantic Script Theory of Humor (SSTH) and its revision by Attardo and Raskin, the General Theory of Verbal Humor (GTVH). Although some aspects of these will have to be outlined in more detail, the reader is assumed to be familiar with, at least, Raskin (1985) and Attardo and Raskin (1991).

In Raskin (1985), the compatibility of the joke text with two scripts, called overlap, is proposed as the necessary condition and the opposition of these scripts as the sufficient condition for a text to be funny. This corresponds to the observation within the incongruity-resolution theories, that incongruity is a necessary feature of humor, but its resolution—for jokes, in the punch line—is the sufficient element (see above, cf. also Attardo 1997). Raskin summarizes it in the following way:

(8) A text can be characterized as a single-joke-carrying text if both of the [following] conditions [...] are satisfied:
(i) The text is compatible, fully or in part, with two different scripts
(ii) The two scripts with which the text is compatible are opposite (1985: 99).

Script oppositeness is a matter of situational, contextual, or local antonyms. Oppositeness can therefore be defined and detected by the combinatorial rules in the semantic links, or rather 'anti-links.' The most likely kind of link constituting oppositeness is the binary category "real vs. unreal" (Raskin 1985: 113). The lexical entry x in the center of the domain that is evoked as one script is semantically linked to an entry non-x, or y, of the opposite script (and often also the trigger for it). This basic opposition can be an actual vs. non-actual situation, or a normal vs. abnormal state of affairs, or a possible vs. impossible situation. This two-fold hierarchy of real/unreal opposition at the highest level and actual/non-actual, normal/abnormal, and possible/impossible oppositions as instantiations of the real/unreal level, is supplemented

by a third level of most concrete pairs of opposition. Typical pairs of opposite scripts as described at this level are, for example (Raskin 1985: 107): DOCTOR vs. LOVER; SEX vs. IMPOTENCE; WISE vs. FOOLISH.

Having started from script opposition as the only element of analysis, the revised SSTH (Attardo and Raskin 1991) encompasses six knowledge resources (KRs) that are used when a joke is generated, and a tentative hierarchy among them: *script opposition*, the highest KR in the hierarchy, the opposition of two overlapping interpretations of the joke; *logical mechanism*, the faulty, local logic of the joke; *situation*, what the joke is about; *target*, the optional butt of the joke; *narrative strategy*, its narrative genre; and *language*, the actual words used in the joke. The expansion leads to the semantic theory becoming founded in linguistics at large. This theory is the main tool used here and relevant parts of it will be illustrated further where appropriate. In terms of the GTVH, this approach is an extension of the existing discussion on the phonology of the *language* aspect of jokes as it becomes crucial in puns who find their *logical mechanism* at this level as I will discuss in more detail in the next section.

The pivot for script oppositeness are triggers that provide a means to detect the switch from one script to another, taking for granted that the first script must be instantiated. This semantic script-switch trigger corresponds to the punch line of the joke and is often verbalized in, or as, the punch line itself. Raskin identifies two sorts of semantic script-switch triggers that can be found in simple jokes: ambiguity or contradiction; hence his emphasis on the disambiguating capacity for a semantic theory. As joke-telling is a form of non-*bona-fide* communication, ambiguity need not be reduced to one unmarked interpretation only, but other interpretations (scripts) evoked by the ambiguity may be

instantiated. The resulting conflict is not so much a blocking barrier, but the sufficient oppositeness for a text to be funny.

Raskin distinguishes regular ambiguity, e.g. *gentleman* as 'man' vs. 'man of quality' and figurative ambiguity, e.g. *innocence* as 'justice' vs. 'chastity,' often reinforced by an auxiliary trigger; syntactic ambiguity, e.g. *with* heading a prepositional phrase either containing an agent 'hand' or an instrument 'spoon,' and situational ambiguity, where incoming information both triggers the new script and continues the original one. Relevant for the present approach is in particular ambiguity created through sound similarity as it can be found in punning knock-knock jokes like the following example:

(9) Knock Knock. Who's there? Cantaloupe. Cantaloupe who? Can't elope tonight—Dad's got the car. (from Pepicello and Weisberg 1983: 67)
cantaloupe -> can't elope

The next section will focus on this specific kind of trigger to create semantic overlap, the faulty logic of word magic through sound similarity.

2.2 The Illogical Logic of Puns: Cratylism

This section discusses the specific type of logical mechanism of puns. We encountered this humorous paralogic incorporated into the GTVH (Attardo and Raskin 1991) as the logical mechanism (LM) in the previous section. The LM in general is formalized and discussed in more detail in Attardo, Hempelmann, and Di Maio (2002). Its specific manifestation in punning was first discussed at some length in chapter 4 of Attardo (1994), which forms the basis of this section. The present study is crucially informed by the concept of LM, which is the second-highest in the hierarchy of KRs as originally proposed. And while it is also the most problematic one (cf. Ruch, Attardo, and

Raskin 1993), its importance as the second key element after script oppositeness will hopefully emerge from the following discussion.

In a much quoted passage, Coleridge identified the two key elements of poetry as "a human interest and a semblance of truth sufficient to procure for these shadows of imagination *that willing suspension of disbelief* for the moment, which constitutes poetic faith" (1817: 169; my emphasis). Accordingly, I would like to suggest that for a text to be a joke as a specific type of poetry, i.e., aesthetic text, we need not only the oppositeness of incongruity, which may well be what makes it of human interest for us. But to reconcile this incongruity at least playfully, so as to facilitate the suspension Coleridge speaks of for poetic text in general, and to accept a relation between the incongruous concepts, the joke needs the LM to appear to bridge that unbridgeable gap between them.

For the illustration of this paralogic of puns I will use an example by Chaucer, who wrote at a time when spelling conventions were not as fixed as today and provided a more forgiving fool's license to pun in writing than English permits today (see note on page 8 of the present study). The following instance, found in the Summoner's Tale and identified by Baum (1956: 231), is *ars* as either *art* or *arse* and it maps the two semantic scripts of EXCREMENT and NON-EXCREMENT onto each other. Because the word can denote both the concepts "lower back" and "art" it creates an overlap between the contrary scripts:

(10) In ars-metrike shal ther no man fynde,
Biforn this day, of swich a question. (SumT III 2222-23)
ars-metrike: *ars metric* -> *arse metric*

The incongruity of this homophonic pun is that of high/low as non-/excrement, referring both to the farting into the friar's hand searching at the arse of Thomas, and the

arithmetical problem of the division of the fart among the dozen members of the friar's convent, that is, a problem for the art of measuring, a possible third reading, in meaning compatible with the "arithmetic" one, and folk-etymologically related to it, making this a particularly complex pun.

The local logic of puns functions on the basis of obviously erroneous reasoning in two steps: first, sound symbolism as a motivated relationship between a word's meaning and it's sound, and second, that this motivated relationship works across sound similarity between two words. The following paralinguistic syllogism, which necessarily implies the contradictory statement in (12) and is nevertheless assumed to be valid for the sake of punning, summarizes the faulty logic:

(11) If meaning motivates sound,
and sound is identical (similar),
then meaning must be identical (similar).

A formalization of the pun logic base on this rationale makes the particular dilemmatic false reasoning involved clear:

(12) $\forall$ a, b, p, q: f(a)=p $\wedge$ f(b)=q $\wedge$ a$\oplus$b $\Rightarrow$ p q
For all sound sequences a, b, and meanings p, q, if function (f) ("denotation") maps a onto p and b onto q and a is in relation $\oplus$ ("has the same sound") to b, then p is in relation ("has the same meaning") to q.

This line of paralinguistic reasoning is named after Kratylos, a participant in the eponymous Platonic dialogue (Plato 1961), who argues for the natural, motivated, non-arbitrary relationship between sound and meaning (cf. Attardo 1994: 152ff). And although cratylism clearly requires a "momentary suspension of disbelief" like all false logic used as an LM of jokes functioning in non-*bona-fide* communication (cf. Raskin 1985: chapter 4), it is false logic, albeit of such pervasive power that we encounter it in

many contexts that are not at all perceived to be non-*bona-fide*, but in which statements implicitly involving this kind of reasoning are taken at face value.

A more sophisticated discussion of the *iconicity* (cf. Morris 1946) of sound and meaning can be found in Humboldt (1836, 1971),[10] who provides a sufficient background against which all issues relevant for the present discussion can be developed. Humboldt subscribes, if not to the arbitrariness of the connection between sound and concept, then to its general inexplicability, when he admits that "to portray external objects that simultaneously impinge on all the senses and the internal stimulation of the mind merely by impressions upon the ear is to a great extent an inexplicable operation as far as the detail is concerned" (1971: 52). But he nevertheless boldly goes on to discuss three types of non-arbitrary, motivated, iconic types of sound-concept relations:

> 1. The first is the directly imitative concept, in which the tone that a resounding object brings forth is reproduced to the extent that articulated sounds are capable of reproducing unarticulated ones. [...]
>
> 2. The second is the indirectly imitative designation, which shares in a property that is common to the sound and the object. [...] It selects sounds to designate the objects which, inherently and in comparison with others, produce for the ear an impression similar to that of the object upon the soul. For example, "stand," "steady", and "stare" give the impression of fixity; [...]
>
> 3. The third is designation based on phonetic similarity in accordance with the relationship of the concepts to be designated. Words whose meanings closely approach one another become endowed with similar sounds; however, in contrast to the designative bracket just considered, emphasis in this instance is not upon the character inherent in these sounds themselves. (1971: 52-53)

---

[10] The relevant passage is in the chapter on the "Lautsystem der Sprachen: Natur des articulirten Lautes," which is the tenth chapter in the original edition (Humboldt 1836), but the seventh (The Phonetic System of Languages) in the translated and edited one (Humboldt 1971). The discussion of iconicity is embedded into the question how language change—that is, a change of either part of the sign, sound or concept, independent of the other—is possible, when their relation is not arbitrary. For a more complete taxonomy see Marchand (1957, 1958), updated in Marchand (1969: 397-428).

It is important, though, to stress the difference between the concept of iconicity found in those words classified in the quote above, most prominently the onomatopoeia of Humboldt's first type, and sound symbolic systems like cratylism, which claim not just a resemblance of sound to concept, but a conventional connection of unspecified logical nature. That is, among sound symbolic systems in general, a distinction is made between imitative sound symbolism (Humboldt's first class), and conventional sound symbolism (Humboldt's second and third classes), which includes cratylism (cf. Hinton, Nichols, and Ohala 1994).

The *klang* association Humboldt describes in his second category comes closer to cratylism than onomatopoeia, but is different in that the shared sound sequence, called *ideophone* or *phonestheme* or *phonetic intensive* (Bolinger 1965), is believed to carry context-independent meaning. In the example the sound sequence [st] is associated with a certain quality (cf. Bloomfield 1895: 409f). Cratylism, on the other hand, does not depend on established ideophone families sharing a sound sequence, but is dependent only on the assumption of a relation between sound and concept in *any* word, or any two words when it is used for the punning LM.

Another type of iconicity subsumable under this category is illustrated by Köhler's famous *maluma-takete* experiment (1929), where the former word was preferredly chosen by subjects to refer to a drawn figure with round shape, and the latter to a star-shaped figure with pointed edges. But this type of sound symbolism clearly involves iconicity motivated by a synaesthetic effect between "sharp" or "round" sound and meaning, thus cannot be classified as purely conventional sound symbolism. For further discussion of sound symbolism in general see Allan (2001) and the research collected in

Hinton, Nichols, and Ohala (1994), and in particular at the level of ideophones see, e.g., McGregor (1996) and Nuckolls (1996), as well as the essays in Erhard Voeltz and Kilian-Hatz (2001)

The third of Humboldt's categories can be considered inverse cratylism, in that the relationship between concepts is considered to influence their sound similarity. But this is surely fallacious reasoning because the sign does not involve symmetric relations: pigment on a paper can stand for a man, but not vice versa (cf. Sebeok 1976). On the basis of this inverse cratylism, Humboldt subscribes to the possibility of general motivated sound-concept relations without the mediation of sound sequences that relate to meaning synaesthetically as in the first two categories, that is a purely conventional sound symbolism that can take scope over any meaningful unit of language. This is the kind of "pure" cratylism found in punning.

In sum, punning cratylism only indirectly—if crucially—depends on the assumption of a motivated relation between a sound sequence and a concept (cratylism). More importantly, it posits an analogical parallelism, as formalized above in (12): if there is similarity between *two signs* in the realm of sound, then there is similarity between the *two signs* in the realm of meaning. Iconicity does not involve such an analogy (except for the inverse cratylism of Humboldt's third category), but focuses exclusively on the motivated relation between sound and concept of *one sign*, which, of course, can be extended into the parallelism of punning cratylism, but does not necessarily entail it.

At this point it seems necessary to point out a common mistake again: The meaning similarity required for humor and provided by the sound similarity in puns, along the lines of cratylistic logic, does not mean that measuring this sound similarity provides a

metric to measure the meaning similarity. Sound similarity in puns interacts with too many other linguistic factors, most significantly to fulfill the target recovery function, and in particular with semantic factors that are unrelated to the semantic resources required by humor, that is, script overlap and, most significantly the script oppositeness (cf. the next section 2.3). In short, a pun with more sound similarity is not a funnier pun, as is often claimed, for example by Fleischhacker)2002; cf. section 4.5.3).

Cratylism in word magic, the assumption of a non-arbitrary relation between a linguistic sign and its *denotatum* as equation or partial identity or at least binding power, plays a significant and well-documented role most prominently in many religious systems, like the cosmogony of the Gospel according to John (1:1): "In the beginning was the Word, and the Word was with God, and the Word was God."

Cassirer summarizes the mythical importance of the word as follows:

> The original bond between the linguistic and the mythico-religious consciousness is primarily expressed in the fact that all verbal structures appear as *also* mythical entities endowed with certain mythical powers, that the Word, in fact, becomes a sort of primary force in which all being and doing originate. In all mythical cosmogonies, as far back as they can be traced, this supreme position of the Word is found. (1946: 44-45; emphasis in the original)

The magical use of words on the basis of their cratylistic equation with their *denotata* is particularly common in connection with names, which are understood not only as a specific symbol, property, or part of a person or deity, and can not only be used to denote that person, but literally *stand* for the person, evoke it, commonly in prayer and incantation (for more examples on tabooed words and names in particular, see Frazer (1935: chapter VI, 318-418). As Cassirer explains, in another paraphrase of religious cratylism,

> in the religious system of the myth the word which denotes that thought content is not a mere conventional symbol, but is merged with its object in an indissoluble unity. The conscious experience is not merely wedded to the word, but consumed by it. Whatever has been fixed by a name, henceforth is not only real, but Reality. The potential between 'symbol' and 'meaning' is resolved; in place of a more or less adequate 'expression,' we find a relation of identity, of complete congruence between 'image' and 'object,' between the name and the thing. (1946: 58).

From this cratylistic reasoning arise the strong word and name taboos, for if the word *is* what it stands for, the name *is* the person, one must use them extremely cautiously. Freud cites the widespread taboo on mentioning the name of a recently deceased (1950: 54-57). In analogy to cratylistic punning logic, this taboo also applies to words similar in sound to the name of the deceased, as Freud explains:

> This taboo upon names will seem less puzzling if we bear in mind the fact that savages regard a name as an essential part of a man's personality and as an important possession: they treat words in every sense as things. As I have pointed out elsewhere, our own children do the same. They are never ready to accept a similarity between two words as having no meaning; they consistently assume that if two things are called by similar-sounding names this must imply the existence of some deep-lying point of agreement between them. (56)

Not surprisingly, the "elsewhere" referred to in this quote is chapter 4 of Freud's discussion of *Jokes and their Relation to the Unconscious*, which reiterates the false analogy of "savages" and children, committing such cratylistic "mistakes that are laughed at by grown-up people" (1960: 147).

The most famous example of taboo based on word magic is the commandment that "Thou shalt not take the name of the LORD thy God in vain" (Ex 20: 7), but extended versions include such diverse phenomena as swearing, summoning ghosts by saying their name, magic formulae, kabbalah and other exegetic exercises (cf. Hausmann 1974: 21). But also onomatopoeia, as discussed above, and poetry, folk etymology and the

explanatory power of etymology in general (Redfern 1984: 84-87) are more or less loosely based on the cratylistic principle, if necessary extended to operate across sound similarity (cf. Jakobson and Waugh 1979).

Of course, the exact opposite position, the arbitrariness of the linguistic sign in which signifier (sound sequence) is related to signified (mental concept) for speakers not through cratylistic motivation, but systematically and conventionally only within the structure of their particular languages, is the canonical assumption, most famously treated in Saussure (1983: 67-69). And this assumption of arbitrariness is generally valid, witnessed simply by the fact that different languages use different sound sequences for similar concepts, e.g., German *Stuhl* and English *chair*. Yet the opposite assumption of a motivated relationship is operational in important areas of language use, as this section has shown, among which punning takes a prominent position.

2.3 Pseudopunning Wordplay

On the basis of the humor theoretic discussion in this chapter, we are now in a position to reconsider the feebleness—and often plain non-humorousness—of what is called "pun" in general, non-technical use. Raskin claims that the script overlap of these jokes is triggered by the quasi-ambiguity based "on purely phonetical and not semantical relations between words" (1985: 116). I agree that in puns, the ambiguity can be achieved through the phonological overlap as logical mechanism. But, crucially, sound similarity—including the ultimate similarity found in identity—cannot alone create the incongruity required by humor, so that a text lacking the latter will be mere wordplay rather than humor. In short, the phonological overlap can trigger and support the script

overlap, but without the appropriate script constellation of oppositeness, it remains too weak to constitute a joke text as defined in (8) above. I will return to this point in the subsection on "definition puns" below (6.10).

Accordingly, I propose that in addition to the overlap in sound of the segment written as *cantaloupe* (target "can't elope"), there is semantic oppositeness, if of the feeblest kind imaginable, like food vs. sex. Otherwise example (9) would not be a joke, and for those who fail to see the oppositeness, it indeed isn't a joke, but merely wordplay. This is, of course, different for those who do identify the semantic overlap, but find it too feeble. They will identify (9) as a joke, but won't enjoy it. And given that humans are desperately good disambiguators with vast semantic networks available to them as well as excellent pragmatic interpreters, if a mere wordplay is uttered, we seek any kind of semantic oppositeness to be able to handle the phonological (quasi-)ambiguity as humor.

The following table makes these possible constellations clearer. Of particular concern is the difference in rows 1 and 2, which are both commonly identified as "puns," whereas the constellation in row 2 actually lacks the necessary semantic script oppositeness to qualify such a text as humor. Note also the peculiar overlap of semantics and phonology in the logical mechanisms of puns.

Table (2): Pun vs. Wordplay in terms of SO and LM

	semantic opposition SO	*cratylism LM*	*interpretation*	*common name*
1.	present	present	punning joke	"pun"
2.	absent	present	wordplay	"pun"
3.	present	absent	a. non-punning joke b. non-joke ambiguity	
4.	absent	absent	non-joke text	

The following are examples for each row:

(13) 1. *Labia majora*: the curly gates. (60)
curly -> pearly
2. Magnet: To some, it is what you find in a bad apple. (145)
magnet -> maggot
3a. Gobi Desert Canoe Club
(non-punning LM: direct juxtaposition)
3b. The square circle ate five freedoms.
4. In case of an emergency, pull cord.

Notice the semantic script of APPLE triggered by "apple" in (13) 2. which crucially supports the recovery of the target "maggot" as well as the lack of semantic oppositeness between any possible scripts including slots for "maggot" or "magnet." As Stanford puts it:

> Nothing is more futile than the irrelevant pun that is based on only a verbal similarity and brings out no contrast, innuendo, or congruity of meaning. Nothing is a clearer sign of incompetent writing than word-play that distracts rather than concentrates the reader's attention. (1972: 72 quoted in Redfern 1984: 20)

In sum, punning includes "word play," but play with words cannot work on the sound level alone, that is, be mere "Klangspiel" (play with sounds), if it strives to be humor as well. But it must be accompanied by "Sinnspiel" (play with meaning; cf. Hausmann 1972: 20) in order for the pun's weak, cratylistic logic to support the opposite overlapping script constellation that would make it a joke. The near failure of this latter requirement, that is, the belief on the part of a joker that he or she can get away with mere "Klangspiel" is what earns puns their deserved pariah status in the family of jokes. I would like to conclude this section on the last note, duly putting into words the professional distance I maintain from my subject of investigation.

3. PHONETIC DISTANCE IN TERMS OF DISTINCTIVE FEATURES

In the previous chapters, I have outlined the issues that participate in the relation between target and pun in paronomasic punning (chapter 1), as well as provided a humor theoretic framework for the discussion (chapter 2). With the present chapter, the discussion can now turn to its central theme, the phonology of imperfect puns. I will first survey previous approaches to this issue in this chapter, before the subsequent one (4) will introduce a new theoretical basis from which I will approach the phonology of the pun-target similarity.

As mentioned above, most early contributions to the discussion of puns from the field of phonology were applications of the concept of distinctive features as a measure of tolerable dissimilarity between a target and its pun. In this section I will discuss the most prominent representatives of this school in chronological order. Sobkowiak's central question—or obvious variations of it, as not all of the approaches summarized in this section deal exclusively or primarily with puns—underlies all approaches discussed in this section: "What kind of rule underlies the transition from the present form to the juxtaposed form of a heterophonological pun in terms of feature geometry?"[11]

[11] It needs to be noted, that in this study the juxtaposed target is assumed to be the underlying form and basis from which the pun present in the text deviates (cf. section 5.2).

3.1 Vitz and Winkler 1973

An early influential study on judged similarity of words in terms of sound is Vitz and Winkler (1973). A central important contribution of their research is that in order to minimize the phonetic distance between words and maximize the correspondence, correspondents have to be optimally aligned. This can be described metaphorically like sliding the two words along each other until maximal correspondence among their sound segments is achieved (1973: 375f). The following is an extreme example, adapted to standard transcription conventions:

(14) #*****rəleʃn# *relation*
#ʌndərrɪ**tn# *underwritten*

Based on alignment, and including a notion of featural identity developed for their approach, Vitz and Winkler calculate the predicted phonemic distance (PPD) of two words and test it against similarity judgments of subjects in a series of experiments. Interestingly, while the PPD does not fare as well as expected with alignment as the only criterion, the revision accounting for phonemic similarity in terms of the classic feature matrix (Chomsky and Halle 1968) fared no better or even worse. Vitz and Winkler conclude that "relatively little of the variance in the rating of complete words is due to factors existing at a lower or more molecular level than the phoneme" (1976: 386).

Another adjustment to their sound similarity model proved to be more successful. Comparing "phonemic clusters," corresponding to the components of the syllable of the monosyllabic words they examined resulted in a much more accurate prediction of perceived distance. An example for the adjusted PPD, not accounting for feature similarity is reproduced in (15) below. The calculations are based on the following

assumptions: The domain of comparison are the syllabic units onset, nucleus, and coda. For each domain the dissimilarity value is calculated as follows: For a given phoneme pair from the two words compared, only full identity of two phonemes in terms of distinctive features counts as towards the PPD as the value 0 ("no distance"). Any dissimilarity is assigned the value 1. Similarity, for example, in terms of place or manner of articulation, is not taken into account, but counts as non-similar, hence leads to the value 1. The values for dissimilar aligned segment pairs are summed up and the mean for the domain is calculated. The total PPD for a pun-target pair is then the mean of the individual means of the domains, so that the value 0 represents highest, the value 1 lowest predicted distance between the tow words.

(15) pl.æ.nt
gr.æ.nt = (1 + 0 + 0) / 3 = .33
spl.ɪ.*t
*pl.æ.nt = (1/3 + 1 + 1/2) / 3 = .61 (1973: 386)

The domains based on syllable structure ensure that consonant clusters in the onset and coda don't have too much of an influence on the PPD, and that the vowel nucleus has a particularly strong similarity cost attached to it.

Several observations of Vitz and Winkler, culled not only from predicted but also rated similarity, are worth repeating here. Higher similarity is perceived when the following elements are identical or more similar: the onset of the first syllable, stressed syllable positions, i.e., vowel nuclei, and the word rhyme (cf. 1973: 381-3). In sum, Vitz and Winkler's evidence "suggests a serial position curve with a strong recency effect and a moderate primacy effect [...]. An explanation of the common poetic devices of alliteration and rhyme is that they take advantage of this curve of relative importance

within a word" (1973: 387). Also, "vowels or at least certain vowels tend to be weighted more than consonants" (ibid.). Together with the alignment requirement and the weak influence of subphonemic distinctive features, this focus on the syllabic structure of the words that are compared for sound similarity is the main finding of Vitz and Winkler. Such resulting observations will be summarized again at the end of all further research surveyed in this chapter in order to systematize them later. This will provide a set of hypotheses which can be evaluated against each other so that a qualified subset can inform the building of an ordered constraint hierarchy in the OT-inspired framework laid out in chapter 4.

3.2 Zwicky 1976

Zwicky discusses imperfect or semi-rhymes in the lyrics of rock music, drawing examples mostly from *Dylan* and *The Beatles*. This application is relevant here because in semi-rhymes, just like for successful punning, minimal distance in sound seems to be a general constraint on the correspondence relation between two segments intended to be perceived as similar. His results are that these *rock rhymes* cannot be described adequately in terms of classical semi-rhyme concepts, namely *light rhyme, apocopated rhyme, consonance,* and *assonance,* but predominantly use two different kinds of deviation from perfect rhyme (1976: 677):

a. *feature rhyme*, "segments differing minimally in phonological features," e.g., *stop - rock*, and

b. *subsequence rhyme*, "X counts as rhyming with XC, where C is a consonant," e.g., *pass - fast.*

In the 700 tokens of rock rhyme, Zwicky finds the following phonemes, separated into consonant and vowel groups, significantly often contrasted in rhyming relations:

(16)

d - z	14	ɪ - ɛ	19
t - k	10	ʌ - ɔ	10
s - z	9	i - e	9
p - k	8	ʌ - a	8
v - ð	7	u - o	6
t - d	7	ɛ - æ	4
p - t	5	ɛ - e	4
d - v	4	a - ɔ	3
b - d	3	(1976: 691)	
t - f	3		
(1976: 685)			

In contrast to the findings of Vitz and Winkler, for both sets of rhymes he emphatically—and obviously oversimplifyingly—concludes that the distance between the two rhyming segments is one feature "in any plausible set of distinctive features" (1976: 686), respectively "in anybody's feature system" (1976: 691). Additional observations that could become relevant in a discussion of correspondence constraints in puns are that alveolar obstruents are easily deleted in casual speech[12] (1976: 681; cf. Sobkowiak 1991: 78-80) and that syllabicity is violated in only 6 of the 700 examples (cf. 1976: 693).

What must be noted again in this context is that rhyming and punning are crucially different domains, so that we must assume distinct, if related, notions of sound similarity to operate in models to describe them. The main differences are the lack of the specific semantic requirements, an LM involving cratylism and SO, in (semi-)rhyming. In

[12] And Bob Dylan's delivery is clearly underarticulated enough to count as casual.

addition, the presence of both corresponding sound sequences in rhyming eliminates the recovery of a target from the functions to be performed by the sound similarity. The latter means that rhyming is far more of an aesthetic, and thus culturally variable, similarity function than punning as witnessed by the notion of rhyme in Old Irish poetry sketched above (cf. section 1.4; see also subsection 4.4.3; Redfern 1984: 99-100).

3.3 Zwicky and Zwicky 1986

In contrast to Zwicky's study on rock rhymes, his later cooperation with his wife on the issue of sound similarity shares the purview with my approach,[13] namely puns, with two notable exceptions. Zwicky and Zwicky "omit all perfect puns and also those imperfect puns involving stress, word division, languages other than English, speakers indicated as having nonstandard accents, matchings of vowels with consonants, or reversal rather than replacement of segments" (1986: 494). I will not attempt to streamline my corpus to the same degree, but include multiple-word domains here, and reflect the strong effect of stress change and segment reversal to similarity through their high ranking in the constraint hierarchy to be sketched below.

Zwicky and Zwicky sort their 2140 pun instances into the following three contrast categories

1. segment/zero (738 instances):
   The lobe was the original earring aid. (74)
   earring -> hearing

[13] This overlap of purview also extends to the material basis of Zwicky and Zwicky's and the present study. As Fleischhacker (2002: 8n10) notes the Zwickys' corpus seems largely culled from Crosbie (1977), just as Sobkowiak's on which my study is based. The only other primary source listed in Zwicky and Zwicky's references is Monnot (1981).

2. featural
involving vowels (547):
Engraving gives one the satisfaction of scratching the etch
etch -> itch[14] (79)

involving consonants (818):[15]
One day Prince Jacques and his sister, Jill, went for a walk with the court jester. Unfortunately, Jacques fell down and broke his clown – and Jill came tumbling after. (50)
clown -> crown
(Zwicky and Zwicky 1986: 494f)

A major complicating issue described by Zwicky and Zwicky is the fact "that particular segments do not appear equally often in targets and puns" (1986: 495). They call this effect *ousting* and define that "when Y appears as a pun substitute for [the target] X more often than the reverse [...], we shall say Y o u s t s X" (1986: 496). Using the voicing feature in stops as an example, the authors show that in general this ousting effect seems related to the underspecification of unmarked features in that "there is a clear tendency for marked (voiced) to oust unmarked (voiceless)" (1986: 497). The ousting effect, also discussed in detail by Sobkowiak (1991: 108-115), will be incorporated into my study by accounting for the asymmetry of pun-target and target-pun segment relations by directionally assigning different cost to them. For example, crucially assuming that more similar target-pun pairs are more frequent in the corpus, it seems to create less sound dissimilarity to replace a voiceless [t] in the target with a voiced [d] in the pun than vice versa, witnessed by the fact that the former occurs 60 times in Sobkowiak's corpus and the latter only 40 times.

[14] Zwicky and Zwicky marked the target as *etch* and the pun as *itch*. I assume in view of the remainder of their discussion that they intended the corrected reading I have given here.
[15] 738+547+818=2103; I was unable to account for the missing 37 instances.

Mohr and Wang (1968) confirm the difference in correspondence of segments with marked and unmarked features. They speculate that the reason for their observation that "members of a marked category should be more similar to each other than the members of the corresponding unmarked category" (1968: 41) might be that "listeners were apparently reacting more to the presence of this (voicing) cue than to the significant lack of it in the other half of the gap [...]. The subjects attended more to the marked half than to the unmarked half of the gap, attaching more psychological weight to the former" (1968: 42). If this observation that markedness to occur more often in the pun than in targets is correct, it will translate into a higher ranking of identity constraint on marked features as they are implying the recoverability of their counterpart phonemes with the unmarked feature. Other than its implicit importance for similarity, markedness in ousting will not be a target for my analysis. A look at Appendix F will confirm that it is not a major parameter in punning.

While Zwicky and Zwicky find only the ə/Ø correspondence occurring frequently in the segment/zero class (50 instances),[16] there are several significant patterns in the featural categories. For illustrative purposes, I chose the following frequent correspondences (more than 12 instances) from their summary, not taking into account ousting effects, that is, which of the phonemes occurred in the target and which in the pun (1986: 497-500) as significant:

(17)	unvoiced - voiced		peripheral - coronal		two feature (alveolar - velar)	
	t - d	57	n - m	34	t - k	26
	s - z	39	d - b	14		

[16] While this is very likely related to the low ranking constraint on schwa syncope in American English (cf. Hooper 1976), the lack of examples in Zwicky and Zwicky's discussion hinders me from confirming this hypothesis.

k - g	29	t - p	13		
p - b	17				
alveolar - dental		*alveolar - palatal*		*posterior – anterior*	
s - θ	21	s - ʃ	38	p - k	13
manner		*vowels: height*		*vowels: tenseness*	
r - l	38	ɪ - ɛ	40	i - ɪ	27
r - w	21	ɛ - æ	29	e - ɛ	18
w - v	14	i - e	26	o -ɔ	12
s - t	13				
f - p	14				

Zwicky and Zwicky summarize (1985: 500): "For consonants, voicing is clearly the most important; there are 163 examples [...]. Position features are the next most important; coronality and anteriority between them account for 197 examples [...]. For vowels height and tenseness are obviously the most important [..]." For my purpose I will start out by ranking the IDENT constraints for these features low. Overall it is striking, that most of the puns of this study "involved a single feature difference (434 of the 818 consonant examples, 264 of the 547 vowel examples)" (ibid.). Sadly, without their material, I am unable to examine the structural effects, like clusters and position of deviant phonemes in terms of syllables, which are at least as important for the discussion of sound similarity in puns, as will become clear from the discussion in the next section. Therefore, I will have to recreate the analysis in the present study, based on the same material as made available to me by Sobkowiak, whose seminal study I will briefly review in the next section, insofar as it isn't based on assumptions form the studies summarized in the previous sections of this chapter.

3.4 Sobkowiak 1991

Starting from Vitz and Winkler's (1973) approach to predicted phonemic distance (see section 3.1 above), Sobkowiak's discussion of English paronomasia presents a major quantificational application of distinctive feature-based accounts for sound similarity in puns, employing an impressive corpus of 3850 puns, which—as noted in the acknowledgements—he generously made available for the empirical verification of the present study. In this sense, his research must be considered the main empirical foundation of my own discussion as presented here. Not only is my corpus a subset of his,[17] but his transcription of the pun-target pairs will form the starting point for my own phonological approach, beginning with a reassessment of the transcriptions themselves.

Sobkowiak confirms the tendency to transpose word-onsets more than codas, that is, stricter identity constraints on the end of words and relates this effect to Hockett's (1958) "Syllable Structure Hypothesis" which "assumes that syllable onsets will be more free to move as they are structurally less adhesive to the rest of the syllable" (1991: 63). In addition to this observation, he summarizes the following significant tendencies that should be reflected in my constraint hierarchy: For a subset of his corpus different from mine, but affected largely by comparable similarity constraints, he finds that "the sounds which are significantly *more* frequent in spoonerisms all are obstruents and noncontinuants, with four out of six noncoronals [...], thus representing an epitomy of consonantism in terms of sonorance hierarchies" (1991: 65). Lesser information load of

[17] I selected 1182 puns that I could analyze for their syntactic and, partly, semantic aspects because their context was available to me from Crosbie (1977).

vowels "means that to maximize understandability of a pun it is reasonable to keep the consonantal skeleton relatively intact" (1991:113).

Sobkowiak also employs a more sophisticated conceptualization of phonemic distance based on the sum of distinctive feature (DF) differences after alignment. Despite its simpler structure, the Vitz and Winkler (1973) approach is supported by Sobkowiak's prediction algorithm. Both confirm that "[p]uns are more similar phonologically than malapropisms not only phoneme-wise, but also DF-wise" (1991: 96).

3.5 Summary

The following observations can be extracted from the survey on distinctive feature-based approaches to target-pun similarity in paronomasic punning. Since word onsets, i.e., the onsets of the first syllable and to a lesser degree word codas, i.e. the consonantal coda of the final syllable, or the rhyme including the nucleus of that syllable, play a crucial role, a more detailed analysis of imperfect puns should include the syllable as a level of abstraction that the model of similarity accounts for. Already in Vitz and Winkler (1973) alignment was based on the notion of syllable and Zwicky (1976) found only six violations of syllable structure in his 700-sample corpus. But, as noted, the syllabic skeleton is more important as tier of analysis to provide a meaningful localization of dissimilar segments. For example, the contrasting of the consonants *d* -> *t*, will be of different significance in the onset of the first syllable of the pun-target pair than in the final coda, where it could be an external *sandhi*—the voicing interacting with that of the onset of the next word—and is assumed to generally have less impact on perceived similarity. In general, the environment of the phoneme(s) in which a pun-target pair

differs, will be an important determiner of the perceived similarity, as it has a crucial influence on its perceptibility as I will discuss in more detail in the next chapter. These two levels of analysis will have to be included in the next evolution of research on imperfect puns as represented by the present study.

Most earlier approaches make strong claims based on the difference between the vast majority of target-pun pairs differing in no more than a single distinctive feature of a single phoneme. Sadly, a brief glance at any collection of imperfect puns will confirm that the picture isn't quite that neat. In particular where extraphonological support for the recovery of the target is available (see chapter 6 below), the pun can differ considerable from it, while the pair is still successfully playing its word magic (see section 2.2 above). Nevertheless, the distinctive feature is still an indispensable tool at useful level of abstraction for the analysis of heterophonic punning. As reformulated and hierarchicalized in feature geometry the distinctive feature, together with the syllable, will form the backbone of the phonological side of my own approach. But I will be able to return to this central strand in the present study— centrally informed by a reformulation of Sobkowiak's (1991) discussion and corpus and incorporating the findings of the other distinctive feature-based studies summarized in this chapter—after an introduction of OT in general and recent work on output-output correspondence constraints in particular in the next chapter.

4. TARGET-PUN DISSIMILARITY AS OUTPUT-OUTPUT CONSTRAINT VIOLATIONS

This chapter will first introduce Optimality Theory (OT) in its standard form as it has been introduced by Prince and Smolensky (1993). On that basis, it will then discuss correspondence in OT, and output-output correspondence in particular. This will form the fundament for the central discussion of the output-output correspondence of target-pun pairs in an OT-inspired model, both in terms of previous studies as surveyed towards the end of this chapter, and for the purpose of the present research. For the following chapter (5) we will then be in a position to reformulate our theory, based on humor theory (chapter 2) as well as the findings for pun phonology in terms of distinctive features (chapter 3), and their recasting into optimality-theoretical terms presented in this chapter, before the theoretical pudding thus concocted can be shown to be edible, that is, when the theory meets the corpus data and can be adjusted and its model enlarged (chapter 6) towards its automatic application to novel input (chapter 7).

4.1 Optimality Theory

OT is a relatively recent and theory in phonology that currently dominates much of the discussion in this subfield of linguistics and has spilled over into related fields like morphology and syntax. It has been proposed in contrast to standard phonological theories to overcome a number of weaknesses of the latter. Most previous phonological

theories, in analogy to the dominant generative paradigm in syntax, model the problems they are created to account for in terms of the structural description of an input and rules specifying the structural changes the input undergoes to become the actual output. In short, they propose three key parts in their models: an underlying structure, rules that operate on it, and a well-formed, "grammatical" surface output. This has the advantages of providing a tractable model for the language issues that have to be described, by providing a finite number of underlying structures and a finite number of possible changes to account for the infinitely diverse output observable in human language. This output is understood as *derived from* the input and rules, or the rules are understood to *transform* the input into becoming the output, hence *derivational* or *transformational* theories. A central problem of such theories are the rules, which are often ad-hoc, non-universal, attempts to explain a language issue and, if more than one rule applies, they need to be ordered.

OT, on the other hand, is a theory of language that does away with rules and replaces them with a finite number of universal constraints that operate on outputs to determine their well-formedness. These potentially conflicting constraints apply simultaneously to outputs, so they don't need to be ordered according to when they apply one after the other, which is an undesirable theory-driven component of generative models, but ordered according to which one has more influence than the others for a given phenomenon in an given language to account for language variability, and which may be reranked to facilitate language change. And in contrast to the rules of generative models, these constraints are soft, violable. In sum, OT accounts better for language universals and language variation then generative models and uses less ad-hoc theory-driven

components in its model. Another advantage of modified OT is its ability to be used for analysis at the high level of formalization required by theoretical and computational linguistics. I will return to this point below. For analysis at this level, generative models are useless, since they can—as their name suggests—only generate output, possibly infinitely, but the rules cannot be reversed to determine underlying input.

Let us look at the model of OT in more detail. It assumes an unspecified generative component (GEN), which creates a set of output forms, the surface representations (SRs) on the basis of an input form, the underlying representation (UR). For example, the UR /in-/[18] of a common prefix in English produces, among all other output forms, the set of output candidates [in-], [im-], [i-], [iŋ-], etc. And, most importantly, OT has an evaluation component (EVAL) that for a given context selects the most optimal of these forms and disqualifies the others as violating higher-ranked constraints, rather than generating that one form directly from the input and a set of rules as generative approaches do.

In sum, OT makes use of four components (cf. Golston (1995: 5)):

(18) UR an unprosodified input string of segments;
SR (candidate) a prosodified output string of segments;
GEN(erator) a function taking one UR to multiple SRs;
EVAL(uation) a function to select the optimal candidate using a tableau of
Constraints, a universal set of ranked, violable well-formedness conditions, the ranking of which is language-dependent; and
Violations, indicators of ill-formedness in terms of constraints.

The main tool used in OT are tableaux with the UR and (selected) SRs in the rows of the left-most column and constraints in the remaining columns, with the lowest-ranking, a violation of which is least damaging to an SR, farthest to the right (see (19) below). The

[18] Notice that the form of the UR here is not claimed to contain a nasal underspecified for place (/N/), because that would prevent an evaluation on the basis of an input-output identity constraint.

ranking of constraints is indicated by solid lines between their columns, while constraints that are not ranked with respect to each other have dotted lines between them. In the cells violations are indicated by an asterisk in the column of the constraint violated and the row of the candidate that violates it.

For example (cf. Pulleyblank 1996: 61-65), in the following tableau the SR [impossible] is not generated through an assimilation rule from the UR /iN-possible/, but selected through weeding out the SRs that violate a ranked group of constraints. The candidates [in-possible] and [iŋ-possible] are excluded through the violation of a syntagmatic constraint on cluster identity ($\text{IDENT}_{\text{PLACE}}$) ranked higher than the other constraints relevant here. This leaves two candidates to be evaluated, which is achieved through a constraint requiring that every segment of the input has a correspondent in the output, in OT-speak "maximize the output with respect to the input" (MAX_{IO}), which is violated by the SR [i-possible], because it has one less segment. Crucially, both of these constraints are ranked higher than a constraint on input-output identity of the features of segments [$\text{IDENT}_{\text{IO-place}}$] that is violated by all candidates but [in-possible].

(19)

/in-possible/	$\text{IDENT}_{\text{PLACE}}$	MAX_{IO}	$\text{IDENT}_{\text{IO-PLACE}}$
☞ [im-possible]			*
[i-possible]		*!	*
[in-possible]	*!		
[iŋ-possible]	*!		*

Three additional elements in the tableau need to be mentioned here: The shading of the right-most column indicated that its constraint is not instrumental to the choice of candidates. It is included to make a point about its ranking in relation to the crucial

constraints to its left. The small hand marks the most optimal form for expository purposes. The exclamation points show for each candidate which constraint violation has excluded it as a possible optimal form.

As this example shows, OT does not take an input and applies rules to it, possibly creating intermediate representations, whose relation to the UR is problematic. Rather it evaluates a candidate set of SRs against one UR, and, importantly, vice versa: Not only is output assessed in relation to a hypothetical input, but the input is evaluated against observed—and unobserved—input, whereas rule-based theories are strictly unidirectional, from input to output. For this very reason, they cannot be applied for automatized analysis of language, that is, output, but only for generation on the basis of a known input.

Because of this bidirectionality of the EVAL of OT, it has also been very successfully applied not only to evaluate output candidates against their input form, but in the form of correspondence theory (McCarthy and Prince 1995) to the relation between related output forms, accounting for their mutual relation, for example, in reduplication and semi-rhymes, as well as in an initial study on the onsets of imperfect puns and proper names.

Another reason it lends itself to the analysis of the similarity of one form to a set of candidates is, of course the non-binary gradational nature of EVAL, namely that we cannot just assume identity or non-identity in imperfect puns, as we can for perfect puns, but there will always be violations of full identity. And, as this brief introduction was intended to show, the evaluation of types and numbers of violations is exactly what OT is good at.

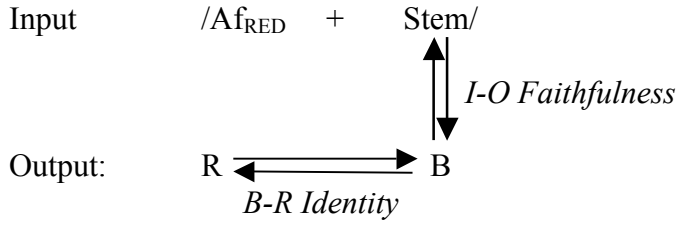
4.2 Correspondence (McCarthy and Prince 1995)

In general, correspondence in OT is any relation between morphologically related words as proposed in the following straightforward definition by McCarthy and Prince:

(20) Correspondence:
Given two strings S_1 and S_2, **correspondence** is a relation R from the elements of S_1 to those of S_2. Elements $\alpha \in S_1$ and $\beta \in S_2$ are referred to as **correspondents** of one another when $\alpha R \beta$. (1995: 14)

McCarthy and Prince's discussion of reduplication is one of the earliest expansions of OT to include not only constraints between input and output forms, but to focus on the relation of morphologically related output forms in terms of correspondence constraints embedded into the OT framework. Their full model includes correspondence between the stem and reduplicant, but the following graph is a rendition of their "Basic Model" of correspondence only between the base and a reduplicant, which sufficiently illustrates the issues relevant for the present discussion, namely Base-Reduplicant (BR) Identity, as well as between the base and the stem, namely, Input-Output (IO) Faithfulness (McCarthy and Prince 1995: 4):

(21)



(Af_{RED} = reduplicating affix; R = reduplicant; B = base)

The main types of correspondence constraints between different types of output are corresponding to the basic constraints between input and output (cf. McCarthy and Prince 1995: 16):

(22) MAX constraints:
Every element of S_1 has a correspondent in S_2. (i.e., no deletion)

(23) DEP constraints:
Every element of S_2 has a correspondent in S_1. (i.e., no insertion)

(24) IDENT(F) constraints:
Let α be an element in S_1 and β be any correspondent of α in S_2. If α is [αF], then β is [αF]. (i.e., correspondent elements are identical in feature F).

Obviously, I expand the notion of correspondence to phonological processes not motivated morphologically, but cratylistically as outlined above in section 2.3. The following subsections will show that such an extension of OT is not without precedent by first summarizing Benua's discussion of morphological Output-Output (OO) relations other than reduplication, then discussing at length Steriade's approach to similarity in semi-rhymes, where identity is usually not motivated morphologically, but largely phonetically and semantically, and finally establishing the basis for my own discussion by outlining Fleischhacker's application of OO constraints in OT to the very issue at hand in this paper, heterophonic puns, as well as a similar extension for the domain of proper names (Lutz and Greene 2001).

4.3 Output-output Correspondence (Benua 1997)

Benua's rationale is to extend OO correspondence to other morphologically related forms than those introduced originally by McCarthy and Prince (1995). She discusses constraints and their ranking in order to explain how "a derived word can mimic its base in features, segmentism, and prosody" (1997: 8), for example, derivational relations like Gr. [illegible] vs. [illegible], or inflectional ones like *house* vs. *houses* in English. Her term for these relations is *transderivational identity*: "Constraints defined over

transderivational (output-to-output or OO) correspondence relations state identity requirements on pairs of words, or PARADIGMS, constructed by morphological derivation" (1997: 3; emphasis in the original).

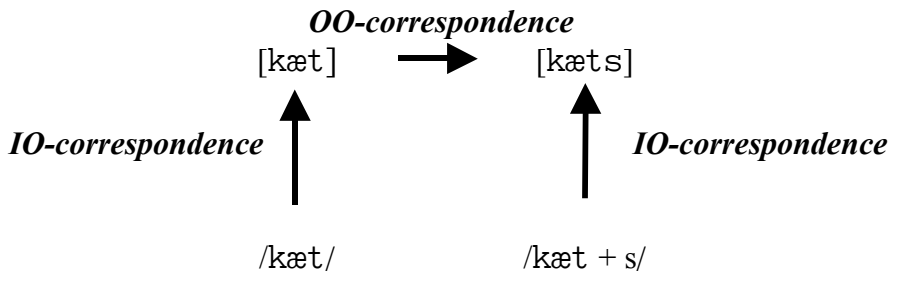
Like for McCarthy and Prince's discussion of reduplication, where "faithfulness constraints on the stem domain always dominate those on the affixal domains" (1997: 4), Benua's ordering of the recursions reflects the paradigmatic relation of the two words between which the correspondence holds. "Paradigmatic relations are asymmetrical, in that the base can influence the derived word, but the derived word never influences the base" (Benua 1997: 53).[19] From an analytical perspective, the same relation holds in the paradigm between a pun and the target it aims at. The target corresponds to the base in a base-reduplicant relationship, in that it is the pivot from which the pun can deviate, but not vice-versa. That is, the target cannot be subject to any change, or in non-generative terms, constrained by correspondence to the pun. But even for puns and targets one must caution that, with a view to pun generation, not any target can form the basis of punning for purely phonological purposes like its uniqueness in a language, thus leading to an—admittedly indirect—influence of pun on target.

Benua illustrates OO correspondence in the diagram reproduced in example (25) with a fairly unproblematic morphologically motivated relation (1997: 7). Of course, she has not summoned this rather sophisticated tool for the analysis of such simple matters, but the complex discussion involving over- and underapplication and related issues is not of concern for my approach to heterophonic punning here.

---

[19] While this is generally true for puns and their target, Benua should qualify the general statement to account for phenomena like paradigmatic leveling as in Latin rhotacism: L *honos, honoris* > *honor, honoris* (cf. Kenstowicz 1994b: 24-26). I am grateful to Mary Niepokuj who pointed this out to me.

(25)



For this kind of morphologically motivated OO-correspondence, Benua introduces recursiveness in the evaluation metric, because both forms between which the correspondence holds must be evaluated against other such pairs, not just one SR against one UR at a time. The following is an example of such a recursive evaluation tableau (1997: 34). It uses the hypocoristic-truncation she focuses on and evaluates the pair in d. with underapplication of the backing of a low vowel before tautosyllabic *r* as least violating:

(26) Recursion (A) Recursion (B)

/læri/	OO-ID	*ær]$_{\square}$	IO-ID	>>	/læri-TRUNC/	OO-ID	*ær]$_{\square}$	IO-ID
a. la.ri			*!		a'. lar			*
b. la.ri			*!		b'. lar	*	*	
c. læ.ri					c'. lar	*!		*
d. ☞ læ.ri					d'. ☞ lær		*	

In this tableau, the SRs læ.ri and lær, and their URs, which turn out to be the optimal forms, are simultaneously evaluated against a candidate set of 2 SRs and 2 URs per candidate. Both the full name and the truncated form are evaluated with the same constraint hierarchy, in which two constraint regulate the individual UR-SR relation, an input-output identity constraint IO-ID, a phonotactic constraint on syllable well-

formedness *ær]$_{\square}$[20]; and the highest-ranked OO-ID constraint, which demands output-output identity between the two SRs.

The key to recursive evaluation is the ranking of the constraints for the base form above those of the truncated form, which "makes underapplication possible without contravening the *ær]$_{\square}$ >> IO-IDENT[BK] ranking. The lower-ranked constraint, IO-ID, compels violation of the higher-ranked one, OO- ID, only because IO-IDENT[BK] is violated in a dominant recursion of constraints" (1997: 49).[21]

The purpose of this section was to introduce the concept of OO-correspondence that is not morphologically motivated. On the basis of this discussion we can now turn to the specific case of non-morphological correspondence between the target and pun in punning. Fortunately, in contrast to the complex interaction of IO and OO constraints described in Benua's model, the present study will not have to pay attention to IO relations, because the corresponding SRs, target and pun, have to be English words, regardless of their IO optimality status. So the cratylistic sound similarity correspondence model does not have to interact with morphological or non-punning transderivational correspondence.

---

[20] Although Benua admits that this constraint is a "brute-convenience" (1997: 33n), this does not diminish its ugliness as an *ad-hoc* solution. I'm grateful to Mary Niepokuj who pointed my attention to this.

[21] An important theoretical question poses itself here: Do we not rather have cyclicity of the same hierarchy of IO constraints (in two cycles in this example), but not really OO-constraints? While the answer to this question lies beyond the scope of my approach at this stage, it needs to be addressed by those focusing on morphology-based OO correspondence.

4.4 Perceived Similarity and Correspondence Environments (Steriade 2001)

Steriade (2001) aims to explicate an important issue that is left implicit in OO correspondence hierarchies and has not been addressed as such before. Based on the common-sensical assumption that OO correspondence is restricted to a large extent by what is perceived as similar sounding, she proposes that a mechanism is required that relates the hierarchies to perceived similarity. In other words, we require the least necessary difference between a pair of output forms, for example, in a semi-rhyme, just as between a paronomasic target and its pun, but we have so far no measure for this difference to be used to order the constraints. This subsection will discuss such a mechanism in Steriade's *P-map*, before the other implications of her work on semi-rhymes to my approach to imperfect punning are outlined.

4.4.1 Theoretical Note

Before I will outline these issues, I would like to point out a different understanding of OT and Steriade's approach on my part: I have assumed perceived similarity to be implicit in all OO constraint hierarchies, in the sense that higher-ranking constraints rule out "more deviant," "less similar," output forms. It is obvious that I am entangled in a metaphorical understanding of a theory that is very metaphorical itself, viz., "higher," "lower ranking," "violate." etc. Accordingly, we must assume that Steriade's insights have made explicit what has been implicit all along. While this may seem to be a moot point, it entails a reason for my disagreement with Steriade who claims that this formerly implicit—now explicit—hypothesis can help to evaluate OT other than making one of its facets tangible. In other words, I claim that it does not add a new external criterion that

enables the empirical verification of OT, but elaborates an assumption on the relation (perceived similarity) between observed material (observable output) and theoretical description (constraint hierarchy).

Steriade seems to have gone the full circle from performance to competence to performance, when she claims: "the P-map is a set of statements about the distinctiveness of contrasts, whereas current correspondence constraints refer to contrast, if at all, indirectly: It is therefore necessary to make explicit the contrast-based nature of correspondence statements" (2001: 27). There is another way to put this circularity: When we observe one output and not another, it is essentially the same thing to say a) that in a description in OT the constraints relevant in the evaluation of this form are in a certain order, implying different levels of perceived similarity, and b) to say that there are different levels of perceived similarity, implying that in a description in OT the constraints relevant in the evaluation of this form are in a certain order. What is gained, and I don't mean to cast a doubt on the fact that this is an important gain indeed, is a higher descriptive adequacy for a model based on a theory all parts of which are explicit.

And it should be noted that, obviously, the very assumption that lead Steriade to make the issue at hand explicit was feeding my hope at the outset of the present project, namely, that OT could help shed light on heterophonic puns, an issue that has all along been identified as crucially involving different degrees of perceived similarity. Steriade started from the theory and found an application, I started from an application and found her (elaboration of) a theory.

4.4.2 The P-map

As stated above, Steriade's work in relation to sound similarity is motivated by the lack of justification of constraint hierarchies in view of perceptual differences of the evaluated forms: "what is needed is a mechanism that relates rankings between correspondence constraints to perceived difference of similarity degree" (2001: 4). Her proposal centers around the concept of a P-map, a map of perceptibility of segmental similarity in relation to different contexts. She defines: "The P-map hypothesis is the claim that one aspect of linguistic knowledge, namely knowledge of similarity, controls grammatical structure, by projecting correspondence constraints and determining their rankings" (Steriade 2001: 13).

The following table is a fragment of such a P-map of Romanian, taken from the handout for a presentation by Steriade and Zhang (2001):

Table (3): P-Map

voicing	V_V	V_#	N_V	N_#
p/ b	p/ b	p/ b	P/ B	p/ b
t/ d	t/ d	t/ d	t/ d	t/ d
k/ g	k/ g	k/ g	k/ g	k/ g
s/ z	s/ z	s/ z	s/ z	s/ z

(size of letter signals relative degree of discriminability: smaller = more confusable)

voice discriminability varies with context (V=vowel, N=nasal, #=word boundary)

The main idea expressed by (20) is that in Romanian voicing contrast is most perceptible in intervocalic environment and least perceptible after a nasal, especially after a nasal and in word-final position. In general, more discriminable segment contrasts are

related to higher ranking correspondence constraints. So perceptual similarity factors are dependent on distance between the segments, contextual cues, and cue duration, an issue I will not discuss here, although it has bearing on puns, albeit at a richer level of analysis and consequently at such a high cost that it is not helpful to enlighten the issue here. These factors determine the ranking of discriminability which in turn determines the ranking of the constraints in a hierarchy.

The P-map aims to explicate how OT was designed to do well what it does, namely to capture the interaction of phonological segments at various levels that do not belong to purely segmental or featural ones. In contrast to such approaches discussed above, the context of the feature-bearing segments ca be taken into account in OT. This is obviously a step in the right direction. Fleischhacker has taken this approach even further in her discussion of onset transfer as we will see below, blazing the trail first traveled by Steriade, and now open for my attempt at a more exhaustive investigation-*qua*-application of sound similarity beyond the level of abstraction of isolated segments.

4.4.3 Correspondence Domains: Rhymes and Puns

The P-map within OT has been developed on semi-rhymes (Steriade and Zhang 2001), a central concept for which is the rhyming domain (RD), circumscribing a string containing one stressed vowel at the end of a line of text in a poem and all segments that follow it. A perfect rhyme is characterized by full identity of this domain in the two segments that form the rhyming pair, identity of the segments before the RD being considered ugly, e.g., *collate-late*, where the onset [l] is outside of the RD, namely just prior to it, but identical. A semi-rhyming pair is assumed to have maximally possible

identity, full identity being precluded here, because different meaning has to be denoted by the two words to which the segments belong (see also the discussion of Zwicky (1976) above).[22]

The central effect Steriade and Zhang examine is that "for some feature F, SR's [semi-rhymes] containing different F values are more frequent in contexts of reduced perceptibility for F." (2001: n.p.). The interpretation for this is, of course, that "judgment of phonological similarity is affected by perceptibility" (ibid.), in contrast to articulatory characteristics of phonemes.[23] And perceptibility is influenced by the context of a given segment, for example, through consonance. Accordingly, Steriade and Zhang not only postulate the usual OO correspondence constraints to hold between the corresponding RDs, e.g., MAX, IDENT[F], but also highlight contextual constraints pertaining to each of the subsegments of the RD and its correspondent.

In puns, the correspondence domain is, of course, potentially larger than in rhymes, because of the constraint on full identity in noted above and the additional metric constraints in rhyming. Generally, the segment comprising the target-pun pair has to be at least one full lexical item that can denote a concept, two of which are required for the humor of the pun to function semantically, with the additional requirements on the scripts

---

[22] Interestingly, at this point the discussion appears to overlap with puns that could be used for the rhyming pairs, a paronomasic pun in semi-rhymes and a homonymic pun in perfect rhymes. Of course, rhymes also require a difference just before the rhyming domain, an issue not specifically mentioned in Steriade and Zhang. But obviously, the following is a particularly infelicitous rhyme (and, of course, atrocious in terms of an attempt at poetry):

> I brought my check to the local bank,
> which was lying on the river's bank.

[23] It has been argued, that similarity perception between words is influenced by the articulatory information stored in their lexical entries, which leads to competition in lexical access (cf., e.g., Cutler and Clifton 1999). To take the analysis to this grainsize is clearly not advisable here.

to which these two concepts belong outline in section 2.1. In short, puns work in general around two complete words represented by the target and its pun.

4.4.4 The C-Constraint: Multiple Survivors

Another central, controversial, and highly valuable feature of Steriade's approach and its application in Steriade and Zhang (2001) deserves discussion in its own subsection: In OT there is only one survivor that is filtered through the constraints. Although there are seemingly better contestants who made it closer to the finish line, it is not a prototypicality or gradational theory, but has binary output, namely the optimal form vs. all other candidates. Assuming the customary ideal speaker-hearer who has only one possible grammatical form for any linguistic choice, OT models this highly idealized unambiguous competence. Of all infinite candidates produced by GEN, only one makes it through EVAL, all the others don't. As we will see, for the discussion of semi-rhymes, Steriade and Zhang (2001) allow for more than one candidate to survive the rigorous EVAL, whose rigor is softened through the introduction of a threshold that lies lower than directly behind the sole survivor in canonical OT.

This main deviation from most other approaches in OT is that they introduce a cut-off point marked by a constraint C in the hierarchy of constraints. Output forms that only violate constraints ranking below that point are allowed to surface, while output forms that violate constraints above the constraint C are filtered out as suboptimal. This is motivated by the fact that not only one, but a number of deviant semi-rhymes are acceptable.

(27) CORRESP.$_1$ >> **C** >> CORRESP.$_2$

↓ ↓

The Unusable The Usable

So the evaluation metric does not just pick the one and only optimal form, which would be the perfect rhyme. Nor does it choose the next best candidate, postulating an additional constraint on full identity that the perfect rhyme or the homophonic pun would violate. So we do not have a form of canonical optimality, but rather an evaluation metric for runner-ups to the best candidate, many of which are not going to be filtered out, namely those that only violate constraints below C.

An interesting question is implicit throughout their discussion, but remains unanswered by Steriade (2001) and Steriade and Zhang (2001). If their evaluation metric is able to distinguish more than one acceptable candidate from the unacceptable candidate(s), is it also a metric evaluating "better" semi-rhymes among the acceptable ones? In other words, could it also be applied as a tool not only to tell puns from non-puns, but also to evaluate more acceptable puns from worse ones?

As should be clear from my discussion in chapter 1, it is my assumption that for the strictly phonological criteria applied the answer would be positive. One must of course be careful, not to make claims about "better" puns on this basis, where better means anything other than closer phonological similarity to the target, in particular any evaluation of the degrees of funniness of puns, as Fleischhacker does when she claims that "there is a positive correlation between pun-target similarity and the goodness of the pun" (2002: 6; see next subsection). Again, from my previous discussion, and even more so from the section on non-phonological factors in heterophonic puns in chapter 6, it

should become clear that even in a form of predominantly verbal humor, such as puns, the semantic element is the most crucial one for the humor of the text.

Yet, the general transferability of the concept of a break-off constraint C to puns should be obvious. In puns we cannot only allow the single phonologically optimal output form to survive, because that would restrict the number of semantically usable candidates too strongly. The likelihood that the one most similar form is usable in semi-rhymes is extremely low, because the rhyming form must be a meaningful string in the language used for rhyming, as well as be an appropriate lexical choice, and compatible with the context. I propose that the likelihood for the single most similar candidate to be a possible target for a pun is even lower, as the semantic requirements on puns are even stricter than those on rhymes: not only must both the pun and the target be meaningful words, but they must be in a humorous relationship of SO (see chapter 2) to each other.

For purely phonological constraints *fire* and *friar* could be considered to stand in a punning relationship, but they are not, unless we find them in a context that conforms to the semantic criteria for a joke, namely evoking two opposite scripts that overlap. This overlap is established through the cratylistic reasoning about the sound similarity (see 2.2), but the *fire* and *friar* are only a target and a pun in the relationship established by all the other elements[24] of this example:

(28) Q: What did the egg in the monastery say?
A: Oh well, out of the frying pan into the friar.

[24] Among these also the contamination through the repeated [fr] onsets in *frying* and *friar* should be noted. But the assumption that *friar* is also a possible homophonic pun on the target *fryer* must be rejected. While *fryer* conforms to the script of COOKING, it would fail to create the overlap to the religious script, that is, the set-up introducing an egg-frying scene in a monastery would be useless. In addition, the force of the proverb "from the frying pan into to fire" must be assumed to force the recovery of the target *fire* from the pun *friar*, and not the target *fryer*.

Without jumping too far ahead in the discussion, let us assume *pyre* to be the most similar sounding English word to *fire*. Letting only this form survive through EVAL, we would not have the possibility to use *friar* as a pun, which would not pose a terrible problem as *pyre* provides us with a word from the same semantic field, an appropriate humorous opposition for which should be fairly easy to find. Using actual words to represent constraint combinations, a hypothetical positioning of C for targets for the pun *fire* could look like this:[25]

(29) *wolpertinger*[26] >> ... >> *fur* >> **C** >> *friar* >> *pyre*

Despite many open theoretical questions, several central ones of which have been outlined in this section, I hope the particular adaptation of OT introduced so far strikes the reader as applicable to possible phonetic distance not only in semi-rhymes, but also in puns, because it makes previously implicit aspects of OT explicit and addresses the central issue of how a certain constraint ordering is motivated. Not only did it inspire me to work on this project, but also Fleischhacker, whose approach to the onset of heterophonic puns has been pointed out to me by Steriade as I was working on an earlier version of this study.

4.5 Onset Correspondence in English Imperfect Puns (Fleischhacker 2002)

4.5.1 Perceptual Similarity vs. Simplification

In many languages, reduplicative processes play a prominent role (cf., e.g., Niepokuj 1997). Usually for morphological reasons, a word contains two similar parts, one of

[25] Of course, *fire* itself is a possible target, and actually the one that should be evaluated as most similar, but the homophonic target, for example as *fire from a job*, is not of concern here.
[26] *Wolpertinger* is the German word for a "jackalope," in which the bodies of several animals are used in a *collage* of taxidermal fancy.

which is the base of which the other, the reduplicant, is a copy. While many complex relations between the two have been described, typically, the reduplicant is simplified with respect to the base. For example, to form the past in this example, the onset of the stem, [sl], is copied and simplified to [s] as part of reduplication:[27]

(30) saíslêp [se-slep] Gothic 'slept' (cf. Braune 1966: §78, Anm. 3)

Despite the tendency to simplification, patterns like the one in example (30) seem to have a privileged status because of their low perceptual difference between base and the reduplicant and are commonly found across languages. Thus, example (30) illustrates the kind of conflicting constraints of similarity vs. simplification typically found in reduplication.

Parallel to this, we have conflicting constraints in punning: On the one hand we require the highest possible similarity for the two purposes of providing a logical mechanism as well as recoverability. On the other hand, we need to have two English words that denote meanings that are compatible with two different scripts. The first requirement is most optimally fulfilled in the case of homophones, but very rarely do two meanings of homophones successfully fulfill the second function. This is where OT-inspired correspondence constraints can be used to model the evaluation of possible candidates that are not fully optimal in order to provide candidate sets that, in turn, can be evaluated by a semantic evaluator that distinguishes jokes from non-jokes. This latter, much more complex task will have to be left for future research (see chapter 7), while the former is out stated goal for the next chapter (5).

[27] The vowel identity between base and reduplicant is incidental as Gothic uses the default vowel [e], spelled <aí>, in the reduplicant irrespective of the stem vowel (Braune 1966: §78, Anm.3; Niepokuj 1997L 144-147).

Fleischhacker (2002) discusses heterophonic puns as evidence for certain privileged patterns of output-output relations in order to support her observations on reduplicative processes. In this subsection, I will outline her discussion on puns in its own right, as it seems to provide a promising attempt to unite the phonological forces at work in punning and, with much extended scope and embedded into the humor-theoretical context provided in the present study, can serve as a starting point for a meaningful account of heterophonic punning.

The variety of onset reduplication patterns encountered by Fleischhacker does not allow for an analysis at a level of generalizations over segment similarity, but leads her to a constraint hierarchy that captures similarity in relation to the environment of the segments as it influences their relative similarity, a level of analysis that I have stressed needs to be included for in punning similarity as well (see P-map above, 4.4.2): "these patterns [...] are sensitive to the relative perceptual similarity of the correspondence strings" (2002: 6). Her main result for our purposes here is that "O_1R_2V -> O_1V is the cluster simplification map resulting in the smallest possible difference between base and reduplicant" (ibid.; O=obstruent, R=resonant, V=vowel). In the following subsections I will follow her discussion of heterophonic puns as it supports this main result and will highlight its relevance for my own approach.

4.5.2 Again: Punning Similarity is more than Acoustic Similarity

Fleischhacker's excellent research goes into as much depth as is warranted by the complexity of the object of her study. It naturally lacks the breadth required by the different purview of a broader approach like the present one. What Fleischhacker

provides for the humor researcher, though, is a template that is extendable to the whole punning domain of a pun-target pair beyond the onset, foremostly to the rhyme of the domain that seems to be a prime target for deviance between pun and target (cf. section 3.1 above) and ultimately to cover relative perceptual similarity in punning in general.

But before I discuss Fleischhacker's approach to English imperfect puns, it is necessary to stress the difference between reduplication and punning, which makes obvious that the exclusive focus on the phonological similarity is a dangerous oversimplification of the relation between the two (see also chapter 5): Whereas in reduplication we see mostly constraints of cluster simplification (markedness, reduplicative constraints) interacting with similarity (correspondence), in punning, the two outputs to be evaluated also have to belong to two opposite, overlapping scripts, possibly even trigger the second script and/or furnish the overlap (markedness, complex semantic constraints) in addition to their sound overlap [=much of the recoverability, all of the cratylism] (correspondence, semantic/syntactic support [idioms, etc.]). But both are "states of compromise reached in the face of conflicting forces" (Fleischhacker 2002: 26).

4.5.3 More Frequent Violations = More Similar Puns?

Three main assumptions underlie Fleischhacker's use of imperfect puns as evidence for her study. They will be evaluated with respect to my own model of imperfect punning in this subsection, and this subsection will also provide the rationale for my use of a corpus. The first hypothesis, that "the pun word must be sufficiently similar to the target that the target can be inferred" (Fleischhacker 2002: 7), is also my assumption for one of

the two functions for the phonological side of this study on punning similarity. But that other factors determine the recoverability at least as much and in interaction with the phonological ones is discussed in a later chapter 6.

Secondly, Fleischhacker assumes that higher frequency in a large corpus is taken to indicate more similar puns (ibid.). This is the only necessary assumption for the present study, and it is not trivially true. The reason that more similar pun-target pairs are more frequent than less similar ones lies in their higher recoverability and, consequently, utility for punning, not in their funniness as I will argue in more detail in the next subsection.

In sum, this reasoning from higher frequency of violation type higher punning similarity is also the basis for the following central claim that we can safely assume that for a given corpus of puns, we more frequently encounter types of similarity constraint violations in target-pun pairs that have a smaller effect on perceived similarity than vice versa. For example, if within consonant clusters in the word onset the final consonant is missing in an overproportional number of puns as compared to the targets in the corpus, this violation is assumed to be less detrimental for the perceived punning similarity than another less frequent violation type. I will use this crucial assumption in chapter 5.

4.5.4 More Similar Puns = Better Puns?

The necessary semantic requirements of humor (chapter 2) that hold for puns, make Fleischhacker's third assumption the most problematic: "there is a positive correlation between pun-target similarity and the goodness of the pun" (2002: 8). This claim can only be salvaged, if similarity is understood to be constituted not only in relation to the phonological environment as Fleischhacker does. Perceived funniness as it can be

modeled within the humor theory used here is largely semantic issue. In view of the necessity of the target not only to be similar to the pun, but also to trigger an opposite, overlapping script, its support for this task from non-phonological factors, and the lack of humor of many pseudopun word-plays found in pun corpora (like Sobkowiak's on which mine is based, Zwicky and Zwicky's corpus used by Fleischhacker is culled from Crosbie (1977)) it is unlikely "that degree of representation in the pun corpus correlates with pun goodness" (Fleischhacker 2002: 10), however one chooses to define "goodness" here. It must suffice to aim at a theory that generates an output that is as similar in composition to that of the observed corpora as possible. More specifically, if a constraint hierarchy evaluates as higher those pun types that are more frequent in the corpus, it has fulfilled its purpose. Of course, to infer similarity from higher frequency is less dangerous than inferring funniness.

4.5.5 Onset Difference in Heterophonic Puns

Fleischhacker selects four types of pun-target pairs differing in their onsets in a fashion compatible to the reduplication patterns which are her main focus, three of which are illustrated in the following table:

Table (4): Three Types of Onset Deviance in Pun-Target Pairs with Examples from Crosbie (1977: 31, 223, 191); O=obstruent, V=vowel, T=stop, S=sibilant, R=resonant (from Fleischhacker 2002: 8)

pun word	**target word**	**pun segment**	**target segment**	**segmental context**	**pun type**	**syllabic context**
Blown-apart	*Bonaparte*	l	Ø	b_o	Ø~R/O_V	medial
surgeon	*sturgeon*	Ø	t	s_ε	Ø~T/S_V	medial
raise	*praise*	Ø	p	#_r	Ø~O/#_R	initial

She continues by calculating the *observed* and *expected* frequencies of the four pun

types: The observed frequency is the ratio of instances of specific segment-type deletions (rows in table (1)) divided by instances of medial deletion of any segment type. The expected frequency is calculated as the ratio of the four pun types to existing word pairs (from the CELEX database) that show the same segment contrast, e.g., *pay* - *play*.

The result of these calculations is first, that the O_1R_2V - O_1V type (first row of table (1)) is significantly prominent, in terms of making up 44.15% of the deletions in the pun corpus (observed frequency), as well as overrepresented, this percentage being 1.34 times higher than the instances of the same segment contrast in a non-punning corpus (expected frequency). Second, the S_1T_2V - T_2V contrast[28] is significantly underrepresented, with an observed frequency of 2.64% which is much lower than its expected frequency of 8.07%.

A major flaw of Fleischhacker's discussion[29] needs to be noted here, namely that she does not take the directionality of the target-pun relation into account. This becomes clear when she describes the relation between the target *Bonaparte* and the pun *Blown-apart* as "O_1R_2V to O_1V" and at the same time the relation between the target *Stabitha* and the pun *Tabitha* as "S_1T_2V-T_2V" (2002: 8). Obviously, if directionality were acknowledged as a parameter the first relation would have to be described as that of O_1V to O_1R_2V, assuming the target as the input and the pun as the evaluated form. Indeed, as modeled in the present study, as well as the previous approaches discussed in chapter 3, the target is assumed to be the basis from which the pun deviates. In analogy to models of reduplication, the target corresponds to the base, and the pun to the reduplicant, a point

[28] Not to be confused with the inverse contrast in row 2 of table (4), where we have the following contrast: T_2V -> S_1T_2V.
[29] I am, again, grateful to Mary Niepokuj who pointed my attention to this issue.

that will be argued in more detail in section 5.2. This asymmetry of the relation has important implications, which I will return to, and these are consequently overlooked by Fleischhacker.

Despite her oversight of directionality in the target-pun relation, Fleischhacker's approach to segment contrast classification seems promising. The next chapter (5) of the present study will present a subset of my pun corpus tagged for punning type similar to Fleischhacker's approach, in terms of contrasting segment(s) and their context, in context in terms of punning domain and syllable position. The pun types thus identified will be tested for their frequency in the remaining corpus, and a constraint hierarchy for puns should be set up to evaluate overrepresented types higher above any cut-off constraint than underrepresented types and exclude unrepresented types as already violating constraints below the cut-off one.

4.5.6 Not CONT, but MAX (and DEP)

For a discussion of reduplication in particular, and possibly output-output constraints in general, and thus also of relevance for pun-target relations, is Fleischhacker's observation that CONTIGUITY constraints are not a promising approach to account for the similarity between forms, be it base and reduplicant or target and pun. CONTIGUITY prohibits deletion and skipping in correspondence relationships, preferring a contiguous copy. Rather, because "correspondence constraints serve the purpose of minimizing the difference between correspondence strings [...] the relevant notion of difference in correspondence relationships is perceptual, not string-based" (Fleischhacker 2002: 12). And in particular, "[t]he perceptual similarity is encoded as the action of *context-sensitive*

Max-C constraints" [my emphasis] in onset transfer in reduplication. The following table summarizes Fleischhacker's constraints enforcing the similarity (2002: 13), where □ stands for perceived difference. We can ignore the constraints that enforce reduplication and simplification (2002: 14, 15) as they play no role in punning:

Table (5): Similarity and Constraints for Fleischhacker's Onsets

similarity scale:	□(C-Ø)/C_V,	□(C-Ø)/#_C	>	□(R-Ø)/S_V	>	□(R-Ø)/T_V
corr. constr.:	MAX_{BR}-C/C_V,	MAX_{BR}-C/#_C	>>	MAX_{BR}-R/S_V	>>	MAX_{BR}-R/T_V

In this table, the first row represents assumed influence on similarity perception of a given correspondence, based on the frequently observed instances in the corpus, and ordered according to this frequency from left to right. The second row translates these into the respective correspondence constraints that are violated. Since to contrast is between a segment in the base and its omission in the reduplicant, these are all MAX_{BR} constraint, e.g., MAX_{BR}-R/S_V: "Don't delete (MAX) a resonant (R) in the reduplicant as compared to the base (BR) after a sibilant (S) and before a vowel (V)."

It needs to be explored whether violations of such MAX ("no deletion") constraints—as well as DEP ("no insertion") violations that are not encountered in reduplication, because it involves simplification (see (30) above), but will be shown to play a significant role in punning—prove to be the more adequate approach to other constellations of similarity in pun-target constraints. In view of the previous discussion it will be clear that my assumption is that they indeed will. Also in this respect, Fleischhacker's method, built on Steriade's extension of OO-correspondence constraints, is a promising basis for an approach to my field of application, where MAX_{PT}-C/x_x constraints—maximizing the identity of pun (P) and target (T) of consonants (C) in

relation to a given environment (x_x)—and similar MAX, DEP, and IDENT constraints will be the main tool to model pun-target similarity.

4.6 Similarity of Personal Names (Lutz and Greene 2001)

Another recent approach to model sound similarity for a specific domain with the help of OT is Lutz and Greene's (2001)[30] study of personal name variants. The type of variation that motivated the application-driven approach of Lutz and Greene is largely based on orthographic variation, e.g., *Conrey* vs. *Connery* or *Geoff* vs. *Jeff*, which occurs when names are transmitted over error-prone audio channels, in particular the telephone or radio, but also due to spelling changes and historical linguistic variation, as well as variation in transliteration from scripts other than Latin, e.g., Russian, Arabic, or Chinese. One major application for name similarity metrics is genealogical research, where comparatively primitive and error prone systems like Soundex (Nara 1995) and Metaphone (Philips 1990) are still in use.

But several differences to the present approach will emerge from my survey of their study, most stemming from the different issues for sound similarity in personal names and some from problems that Lutz and Greene have overlooked. First, a crucial difference in purview exists: pun and target are real words, therefore semantic factors must be taken into account, more difference must be allowed, stricter syllable structures can be expected than in proper names, and different effects of conventionality on perceived punning similarity. Second, and equally importantly, the environment of

[30] Their study is published only as an unpaginated technical report, therefore no page numbers are provided in this subsection.

violations is not taken into account by Lutz and Greene. Third, in target-pun correspondence there is clear directionality of the constraints, the target being the pivot form (parallel to the base in reduplication) from which the pun (parallel to the reduplicant) deviates. In personal name relationships, on the other hand, there is a historical relationship of newer to older forms, implying the older forms as the basis from which the newer ones can be understood to deviate. But, more typically, two current name variants that share an underlying older form are compared, for which directionality could only be useful, if this operation is done in two separate steps of comparing each of the current forms to the shared ancestor and then combining the results. This is, of course, only possible if the ancestor is known. It is clear that Lutz and Greene are operating with a notion of asymmetry when they state the "costs [...] are not inherently reflexive: the cost of [X] going to [Y] will not necessarily be the same as the cost of [Y] going to [X]." But, regrettably, it remains unstated which directionality between names is underlying this asymmetry, and why.

In general, Lutz and Greene's computational approach to name comparison is based on similar—but in essential aspects different—theoretical assumptions, not least its implementation of a significantly modified version of OT. It is therefore useful to outline their concept of phonological similarity in more detail. They base the initial similarity algorithm on pairs of corresponding IPA characters between which they determine edit distance just as in Sobkowiak (1991) and the present study. But, as mentioned, a major difference to the approach here lies in the inclusion of the environment of the segments that do not fully correspond is taken into account. A simple edit distance algorithm would simply assign a value of *1* for every non-matching pair and sum up these values to

produce the total score for a given segment, yielding as imprecise a value as produced by the earliest Vitz and Winkler method (see section 3.1).

Realizing this problem, Lutz and Greene modify their initial algorithm "to score pairs of different characters gradiently according to their phonological relatedness." The different grades were collected into a cost table matching each pair of the phoneme inventory with a value ranging from 0.0 to 1.0, related to the number of distinctive features in which the two phonemes of the pair differed. The key to this modification is, of course, the central problem of modeling the notion of phonological relatedness and how it is formalized for the purpose of creating a cost table matching each phoneme pair to a value.

In an attempt to enhance the similarity performance of the cost table, Lutz and Greene go the same way independently chosen in my approach, namely to use the hierarchy of feature geometry in autosegmental phonology rather than simple feature counts based on SPE,[31] and to base the values conceptually on OT constraint violations (see section 5.7). In this context they note the important difference of conceptualizing constraints in this way with respect to standard OT, a difference I require for identical reasons in my approach: "constraints are used to build an abstract cost model that gives a measure of the costs associated with bridging the feature differences between any two phonemes. A final distance cost for any two phonemes is the sum of the individual costs for each violated constraint. The constraint rankings determine the relative weight of each constraint in the

[31] On the problem of using an articulatory model like feature geometry for similarity judgments, which are largely acoustic (see subsection 5.3.2).

calculations." In other words, the cost table represents the constraint hierarchy in a convenient formularization for a computational implementation.

Let us now turn to the phonological assumptions about sound similarity that underlie Lutz and Greene's model. In general, sonority is considered the key parameter. The following aprioristic set of violations is the core of their cost matrix for changed—in contrast to inserted or removed—segments (corresponding to my IDENT(f) constraints). Accordingly, the most basic root level of the feature hierarchy is

(31) consonantal >> sonorant.[32]

This assigns a higher cost for a change from a C to a V, a correspondence violation that leads to severe loss of similarity, than to those constraints that just changes the sonorance between the segments at hand. These constraints assign varying costs that are related to how distant the corresponding pair of sounds is in terms of the standard sonority scale:

Vowels > Glides > Liquids > Nasals > Obstruents (Fricatives > Affricates > Stops).

The sonority scale implies higher similarity between sounds that are adjacent in it than sounds that are further apart, so that a violation that has a nasal and a fricative as corresponding segments leads to less perceived distance between the respective names than one involving a vowel corresponding to a stop.

It remains unclear from their discussion, how exactly Lutz and Greene would formulate the corresponding correspondence constraints, but we must assume individual constraints, all of which are ranked—that is, practically assign a different cost—for each of the possible relations in the sonority scale. For example, IDENT(son)-GN >>

[32] Lutz and Greene use this more current term rather than "resonant" as implied, for example by Fleischhacker's acronym R for this sound class. I will treat these terms as interchangeable.

IDENT(son)-VG would mean that a violation of identity of the sonority feature that involves a vowel in glide is more costly than such a violation in a corresponding pair of a vowel and a glide.

Lutz and Greene handle epenthesis and deletion, i.e., violations of DEP and MAX constraints, symmetrically in the same hierarchy at the root level and using the same analogy between higher cost and larger sonority difference. In view of the asymmetric ousting effect (see sections 3.4 and 5.2), it seems highly problematic to assign the same cost to insertion and deletion of segments, hence my own independent ranking of DEP and MAX constraints.

The remaining three feature dimensions, place of articulation, manner, and the laryngeal feature are assumed to contribute to similarity distance in terms of the following hierarchy (contrasting with Zwicky and Zwicky's (1986) model; cf. section 3.3):

(32) place >> manner >> laryngeal

Note, again, that these aren't constraint hierarchies, but hierarchies of features in terms of the autosegmental feature tree, in terms of how much they contribute to perceived similarity. But they imply a ranking of constraints against the violation of similarity, such that, for example, a change in manner from one correspondent to the other is less costly than a change in place of articulation.

Among the place features, the following hierarchy is assumed to hold, again, implying a higher cost for larger distances traversed at this level in the feature hierarchy by a given correspondent pair:

(33) labial, dorsal >> coronal >> pharyngeal

In more detail, this hierarchy represents the violation penalties given in (34). Note, that unlike in the previous hierarchies in this section, the following one does not indicate the importance of a feature in terms of its influence on perceived similarity when it is changed, but a set of constraints on these violations ranked according to their cost, and including directionality of the violation (→), respectively the lack thereof (↔).

(34) least costly coronal → pharyngeal
coronal + dorsal → coronal
coronal → coronal + dorsal
labial, dorsal → pharyngeal
labial, dorsal → coronal
pharyngeal → coronal
coronal → labial, dorsal
most costly pharyngeal → labial,[33] dorsal; labial → dorsal[34]

Among the manner features [continuant], [nasal], [strident] and [lateral], Lutz and Greene posit the following discrete hierarchy for [continuant], claiming convincingly that "[n]o inclusive universal constraint hierarchy of any kind is yet, or likely, to be established for the highly discrete manner features." They choose not to account for the other manner features as "[r]elatively small, reflexive costs are incurred for any changes in the[se] simple binary features."

(35) -continuant >> ±continuant >> +continuant

This, again, is based on the sonority scale as manifested in the following hierarchy:

(36) stop >> affricate >> fricative >> sonorant

---

[33] It seems odd that this violation is most costly in either direction, when the change from coronal to pharyngeal is least costly. In addition to Lutz and Greene's assumption of "large physical distance within the supralaryngeal articulatory space between dorsals and labials," the low acoustic quality of pharyngeal features can be assumed to account for this.

[34] Interestingly, as Mary Niepokuj pointed out to me, in previous attempts at acoustical feature systems, [labial] and [dorsal] (or [peripheral] and [medial]) were grouped under [grave], because they share a "concentration of energy in the lower (*vs.* upper) frequencies of the spectrum" (Jakobson and Halle 1956: 31). Judging form the results of Lutz and Greene (2001), the similar frequency, while an acoustic common ground, seems to have less influence on the psychoacoustics of perceived similarity.

In more detail, this hierarchy represents the following constraint violation, most notably the fact that "moving in the direction of a positive value for continuance is less costly than moving toward a negative value for continuance," a notable exception to the general symmetry of Lutz and Greene's approach, but consistent with the notion of lenition as a common process[35]:

(37)	least costly	-continuant → +continuant, ±continuant
		±continuant → +continuant
		±continuant → -continuant
		+continuant → ±continuant
	most costly	+continuant → -continuant

Voicing as the only laryngeal feature relevant in English was assigned little overall cost, again accounting for asymmetry, as well as the markedness ousting observed by Zwicky and Zwicky (1986; cf. section 3.3), "based on the current theory claiming that voicing is a privative feature, and is marked in obstruents:"

(38)	lower cost	`+voice → -voice`
	higher cost	`-voice → +voice`

On the basis of these assumptions, a

> feature table was prepared to itemize the feature values for all phonemes in use in the system. Using this table as input, a utility program was written to evaluate the constraint violations implied, within the model just described, by the pairwise transformation of all phonemes. Costs were calculated for each pair of phonemes and output to a file for input to the system.

As hypothesized, the similarity calculations based on this OT-inspired approach and "more in line with current phonological theory" improved the ranking results compared to those based on the SPE feature-based approach. This contrasts to Vitz and Winkler's approach (see section 3.1), and indicates, somewhat trivially, that the SPE model, as a

[35] Thanks, again, to Mary Niepokuj for making me aware of this.

largely articulatory approach to features, is a less appropriate candidate for the description of sound similarity than more acoustically-based models like the one devised by Lutz and Greene.

Some fine tuning of the cost table that lies at the center of Lutz and Greene's approach had to be done in view of further testing against data. In particular,

> costs for the epenthesis and deletion of the strident coronal fricatives [s] and [z] were lowered to that of epenthesis and deletion of nasals and liquids, rendering them exceptional among the obstruents. Similarly, epenthesis and deletion of the glottals [h] and [/], were adjusted to the same cost as epenthesis and deletion of the glides.

This resulted in the following adjusted scale for epenthesis- (DEP violation) and deletion-like (MAX violation) contrasts:

(39) < lower cost — higher cost >
< higher sonority — lower sonority >
[vowels] [glides, [h], [/]] [liquids, nasals, [s], [z]] [obstruents]

Another important adjustment concerns nasals, who show extensive place assimilation and are perceptionally extremely similar, so that the high cost put on place feature changes in general had to be significantly lowered here.

Overall, Lutz and Greene's research is a major advance over previous sound similarity models, despite the fact that it ignores much of the OT research outlined above. In line with one of their suggestions for future research, the present study is informed by a corpus-based approach, and will take the issues overlooked by Lutz and Greene into account, most prominently the environment of the constraint violations and the motivated asymmetry of most segment relations.

5. PUNNING CORRESPONDENCE CONSTRAINTS

At this point we are in a position to recast the observations of earlier approaches into a formal correspondence theory for target-pun pairs and a tentative hierarchy of OO-inspired correspondence constraints. The previous chapters have outlined the purview of this study (chapter 1), introduced the general humor-theoretic background (chapter 2), surveyed previous models and their hypotheses (chapter 3), and developed a theoretical framework consisting of context-sensitive correspondence constraints modeling perceived similarity judgments over the domain of English punning (chapter 4). I will continue by briefly summarizing the results from chapter 3 insofar as they are able to inform the OT-based hierarchy of my new model. This hierarchy will then be related to the empirical findings of pun-target correspondence violation in the corpus of 1,182 puns, and improved with the help of these findings as they are accommodated by the phonological theory outlined above.

5.1 Overview of Previous Results

The distinctive feature as the main or sole parameter for modeling sound similarity has been found to be inadequate, at least insofar as standard sets of distinctive features are largely articulatory and not acoustic. Perceived sound similarity, on the other hand, insofar as it is non-arbitrary and culturally variable, is obviously chiefly an acoustic issue.

Accordingly, a modified model based mainly on sonority and acoustically biased feature geometry will be used here.

The phoneme as the main subunit of corresponding targets and puns has been confirmed by all studies, but its environment, in particular its position with respect to the syllable and word has so far not been accounted for sufficiently in any of the previous models except for Vitz and Winkler (1973) who grouped phonemes into onset, nucleus, and coda thus effectively introducing the consonant cluster as an intermediate level in their analysis.

Identical assumptions were made about the relative importance of consonants as compared to vowels. All studies assumed vowel quality to be more violable than consonant quality. Within the group of consonants, most frequently stops have been found to resemble other stops most closely so that they are most interchangeable, followed by stops and fricatives sharing the same place of articulation. Another frequent type of correspondence is that between the voiced and unvoiced variants of otherwise identical consonants, indicating that in terms of the classic feature groups, voicing may be different between correspondent consonants with the least cost to similarity, followed by place of articulation, followed by manner with the highest cost. Note, again, that while this hierarchy is the best model available for such distinctive feature classes, these classes themselves are not the best-suited model type for the description of sound similarity. Predictably, vowels tend to be more frequently replaced with vowels that are close to them in height and backness, these terms being metaphors for distinctive features that capture vowel qualities which are more directly acoustically based than those of consonants.

Alignment of the target and pun is not a simple issue, as the number of segments between the two often varies, and the identification of the correspondent segment to a gap is not *a priori* clear, and only to a certain degree predictable from syllabic structure. On the basis of a similarity metric, on the other hand, alignment becomes a simple computation of which choice of correspondent segment to a gap leads to the overall greater similarity between the two words to be aligned.

Syllable structure is only very rarely different between two correspondents that are to be perceived as similar, and within the syllable, consonants in the onset differ from one correspondent to the other more frequently, followed by the vocalic nucleus, and the coda is usually identical, a difference here resulting in low perceived similarity, indicating the importance of the end of words as in rhyming.

In terms of the autosegmental feature-based approach discussed in the last section of the previous chapter (4.6), the following hierarchy of relations should be accommodated by a model of sound similarity,[36] in which > indicates the sonorance relationship on which these hierarchies are based, >> a higher cost in a change of the feature on the left than in the feature on the right, and -> a directional cost difference between target and pun, and <-> a symmetrical, reflexive cost relation:

(40) Vowels > Glides, [h, ʔ] > Liquids, Nasals, [s, z] > Obstruents (Fricatives > Affricates > Stops)
- a. consonantal >> sonorant
- b. manner >> place >> laryngeal
  - 1. labial, dorsal >> coronal >> pharyngeal
  - 2. coronal -> pharyngeal least costly
    - coronal + dorsal -> coronal
    - coronal -> coronal + dorsal

[36] I streamline Lutz and Greene's (2001) model greatly, omitting, for example, the consonantal vs. sonorant contrast which is unattested in my corpus.

	labial, dorsal -> pharyngeal	
	labial, dorsal -> coronal	
	pharyngeal -> coronal	
	coronal -> labial, dorsal	
	pharyngeal -> labial, dorsal;	
	labial <-> dorsal	most costly
3.	-continuant > ±continuant > +continuant	
4.	-continuant -> +/±continuant	least costly
	±continuant -> +continuant	
	±continuant -> -continuant	
	+continuant -> ±continuant	
	+continuant -> -continuant	most costly
5.	+voice -voice	least costly
	-voice +voice	most costly

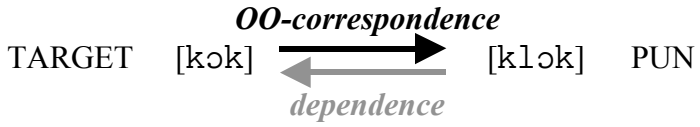
Finally, it is important to note, again, that these possible differences between the correspondents do not occur symmetrically, that is, certain replacements are significantly more frequent in a pun as compared to the target than in the target as compared to the pun. I will return to this issue in the next section.

5.2 Asymmetrical OO-Correspondence in Punning

This subsection will summarize the theory of sound similarity developed in chapter 4, as the basis for a model into which the findings summarized in the previous subsection will be incorporated. The following diagram illustrates the phonological correspondence relations motivated by the cratylistic-semantic relation between the pun and its target (for a punning context, see the example in footnote 3). The relation to canonical OT correspondence theory, especially in its expansion by Benua (1997), is obvious. The main difference to her approach is that "[t]ransderivational OO-correspondence relations are the phonological reflex of a morphological relation between two words" (Benua 1997: 28), but punning OO-correspondence does not require morphological relations between a

pun and its target, but rather mirrors the cratylistic pseudo-semantic relation between them:

(41)



In this figure, the asymmetrical OO-correspondent relation (black) is represented by an arrow pointing from the target to the pun, indicating that the target is the base against which the pun is evaluated (as well as dependent for its recovery on the pun). For example, if there is an additional phoneme in the pun, as in the figure, this difference is conceptualized as analogous to an insertion. Correspondingly, the violation is that of a DEP constraint that forbids the insertion of segments into the pun for which there is no depending segment in the target form. Inversely, the pun is dependent on the target as the pivotal form (grey), being derived from it, in the metaphorical sense of the classic generativist approach to the punning relationship. In sum, the present study assumes the target as the base for correspondence in relation to which differences in the pun will have to be evaluated. As mentioned above, this is analogous to the directionality of correspondence constraints as found in the reduplicant as evaluated against its base in reduplication.

It is tempting to look for a correlation between the meaning relations of a pun and its target, and the phonological OO-correspondence relations. But this fallacy obviously means to take at face value the cratylistic reasoning that uses the faulty logical mechanism underlying punning itself and has been outlined above. Yet, as noted above, neither Fleischhacker shies away from it, nor does Sobkowiak: After quoting Apte's (1987) conjecture that the "degree of ambiguity," a notion the measuring tool for which I

would be delighted to see, must be neither too high nor too low, he claims "[i]t would be interesting to see if, and how, this dynamic equilibrium manifests itself structurally in puns" (1991: 103). He never follows up on this promised investigation, though, and I am sure he would realize the false analogy, based on the pseudologic of cratylism, fairly quickly. It remains to be repeated at this point: I am making no claim about any motivated parallelism in the phonological and semantic relationships in a pun other than the playful one outlined in section 2.1.

5.3 Constraints on Punning OO-Correspondence

This section will present the basic inventory of OO-correspondence constraints of punning based on the model developed in this study and the findings from previous studies. This inventory, put into a tentative hierarchy, will form the basis for empirical verification, and corresponding adjustment and refinement of the theoretical assumptions behind the model. First, the main constraint types on correspondence will briefly be outlined again.

According to the directionality of punning correspondence discussed in the previous section, constraints are violated in the pun, not the target, while the target is considered the anchor to which the pun must strive to be similar. General IO constraints do not have to be taken into account in any form, because both forms, pun and target, are surface forms that have successfully survived an IO evaluation.

Thus, the constraints that have been found to be relevant for the modeling of punning sound similarity, are all OO correspondence constraints. In order of frequency of their violation and, based on the assumption that more frequent violation implies less

detrimental effect on perceived similarity (cf. 4.5.4), in reverse order of the cost of that violation, these families of constraints are:

(42)	IDENT(f)$_{PT}$	no change at level and type of feature f
	MAX$_{PT}$	no deletion
	DEP$_{PT}$	no insertion
	MAX/DEP$_{PT}$-□	no deletion/insertion of a syllable
	IDENT(‘)$_{PT}$	no change of stress position
	LIN$_{PT}$	no metathesis

IDENT, MAX, and DEP constraints have been discussed in detail in chapter 4 above, and some detail, in particular about the internal hierarchy among the Ident constraints will be added in this section. And I will return to MAX/DEP$_{PT}$-□ and IDENT(‘) as constraints on suprasegmental aspects of identity in section 5.4 below.

Linearity (LIN) is the constraint that prohibits metathesis (McCarthy and Prince 1995). While it is extremely rare—only 7 cases in the whole corpus were found by Sobkowiak—its cost is not assumed to be high, because it may result in very little or very much perceived dissimilarity, depending on the type, position, and environment of the metathesis. The central problem is to identify the size of the domain in which to look for metathesis, as well as to analyze the metathesized segments within the domain in terms of their position in the words and syllables, for example onsets. The problematic nature of the domain size is illustrated by the predilection of metathesis to occur in onsets across word boundaries, while observing morphological segmentation, so-called spoonerisms. Although it has been successfully incorporated into OT-based correspondence theory (Hume 2001), and accordingly could be incorporated into the OT-inspired model here, I have chosen not to implement it, simply because it is too complex and too hard to

discover computationally to make the effort meaningful for the few instances the system would have to handle, as a tool for both analysis and generation.

A crucial decision to be made for constraints from all families is the question at what grain size the constraints' operation should be described so as to achieve descriptive and explanatory adequacy. The actual affected segment will be analyzed at the phonemic level, the environment at the level of natural class. The level at which the distance itself is analyzed is the node in an autosegmental feature tree adapted to account for acoustic rather than articulatory dimensions. I will turn to each of these levels of analysis in the following subsections.

5.3.1 Phoneme Classes

Classes of sounds are considered the most promising level of abstraction along the continuum *allophone - phoneme - class - C/V*. The general assumption about these classes is that "they can be uniquely identified in terms of a single conjunction of features" (Clements and Hume 1995: 245). This traditional assumption, while generally correct, disregards the fact that certain sounds may behave similarly, although they share less distinctive features than other sounds that do behave differently, especially when these features are articulatorily-based. Thus, a better assumption is that "it is much more common for [segments] to be grouped into natural classes because of specific properties relating to the way they are heard" (Iverson 1995: 611). Therefore, these classes should be acoustically-based classes of sounds that are perceived as similar and, consequently, behave similarly in identical environments. This is a criterion for natural classes that can be a significant abstraction for the purpose of analyzing sound similarity. The following

such classes have been adapted from Fleischhacker (2002) and Lutz and Greene (2001) and have been used in the tagging tasks of the present study:

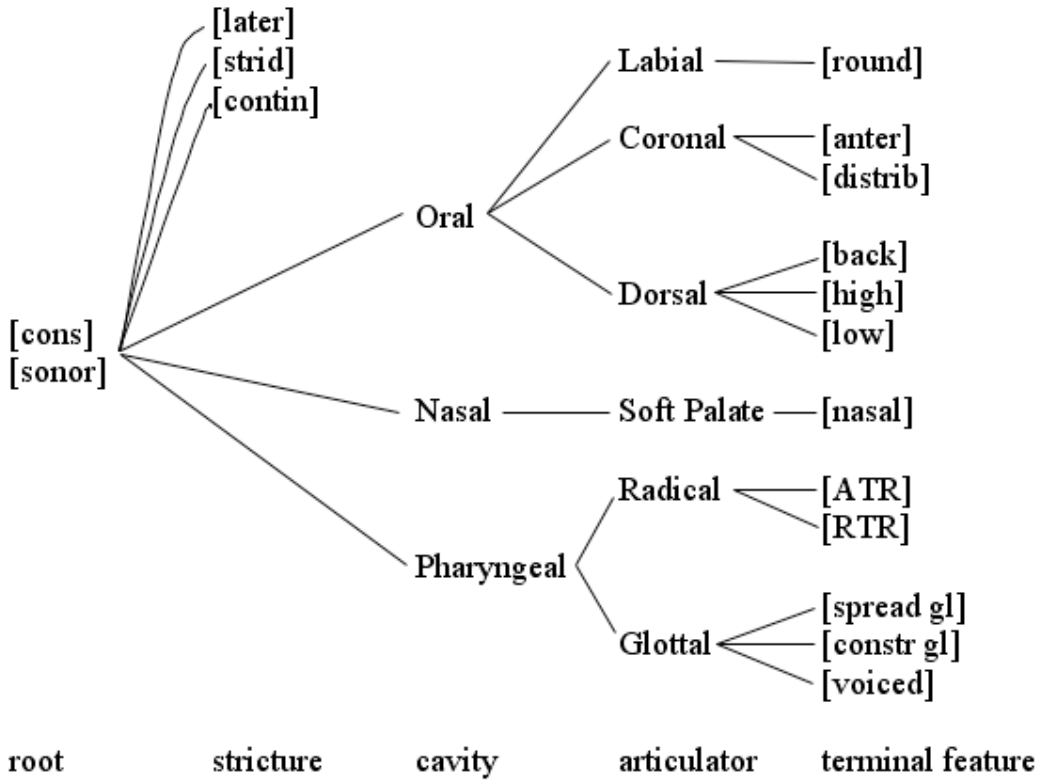
(43)
- T — stop [p, b, t, d, k, g, ʔ]
- N — nasal [m, n, ŋ]
- S — sibilant [s, z, ʃ, ʒ, tʃ, dʒ, θ, ð]
- R — (other) resonant/sonorant [r, l, w, j]
- O — (other) obstruent: fricative [f, v, h]
- V — vowel with additional marking
  - b — back
  - f — front
  - h — high
  - l — low

5.3.2 Acoustic Feature Tree

For the constraint hierarchy, and the implied cost hierarchy of their violation, the level of feature will be used for consonants. The model on which the adaptation of the present study is based is on the feature tree. The motivation underlying the feature tree model is not that linguists cannot think outside of tree metaphors, but the fact that the features of consonants aren't freely combinable. The presence of one can exclude the presence of another. For example, the specification [– consonantal, – sonorant] is "precluded because a [– consonantal] segment's lack of constriction in the oral cavity implies spontaneous voicing indicative of a [+ consonantal] segment" (Kenstowicz 1994: 147). Another reason to use a tree implying a hierarchy is that some features are subdivisions of another feature, for example the articulator features are dependent on their cavity (cf. 44), and only if that higher feature is specified for a sound, do subfeatures of it become relevant, or inversely, the specification of a terminal feature implies the specification of that whole *branch* of the *tree*. A basic tree that captures these feature

groupings and combinatorial restrictions is given in (44), adapted from Kenstowicz (1944: 146; based on Halle 1992).

(44)

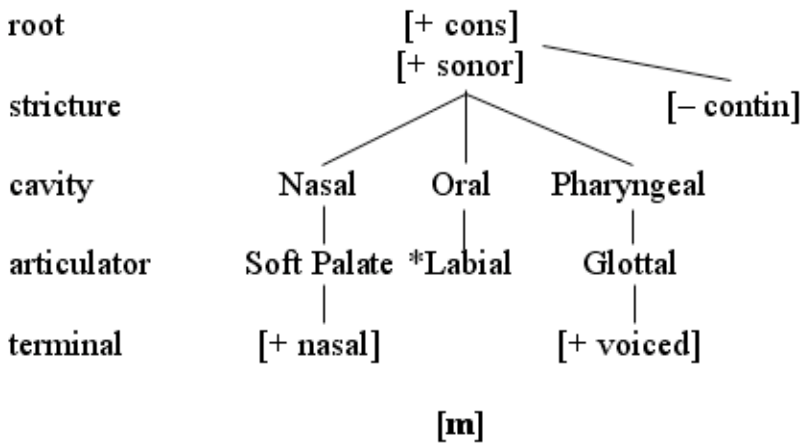


With the help of such a tree any sound con be described by using its root features, [consonantal, sonorant], the more freely combinable stricture features, [lateral], [strident], and [continuant], and the terminal features, [rounded], [anterior], [distributed], [back], [high], [low], [nasal], [advanced tongue root], [retracted tongue root], [spread glottis], [constricted glottis], and [voiced], grouped under articulators, which in turn are grouped according to cavities.

With the help of this tree, we can now describe a segment not only in terms of their features, but at the same time indicate relations among the features of a segment, and, more importantly in the context of autosegmental phonology for which such trees were

invented as a model, but not relevant for the adaptation of the model here, also the specification and interaction of features across segments. Example (45) shows how [m] is represented. The asterisk specifies the major articulator realizing the root features (from Kenstowicz 194: 149):

(45)



After this brief introduction of the articulatory feature tree, we can now turn to the discussion of IDENT(f) constraints for consonants, which will be represented by the lowest point in an adapted acoustics-based feature geometry tree at which the segments in question are identical, or, in other words, the point of most identity. For this purpose, we need a stricter notion of hierarchy than in (44) to actually have such a point, as well as further streamlining of the higher level groupings of terminal features in terms of the acoustic natural classes proposed in (43). The proposal in (46) is based on the assumptions that a violation of manner of articulation identity of a consonantal segment has a stronger effect on perceived similarity than a violation of place of articulation or nasality, both of which incur similarly lower costs. The least effect on perceived similarity is assumed to be achieved by a change of voicing. (46) is necessarily a generalized representation inferior to the more detailed

(46) **root** root
- **man** manner
  - app approximant
  - fri fricative
  - sto stop
  - **plac** place
    - cor coronal
    - lab labial
    - dor dorsal
    - rad radical
  - **nas** nasality
    - **voic** voice

The tree in (46) is necessarily a simplified representation as compared to the more detailed—and directional—account in (40). But as an overview it illustrates the general assumption well. The indentation represents the following hierarchy of intermediate levels in terms of their importance for perceived similarity (cf. Kenstowicz 1994a: 452ff, Halle 1992):

(47) root >> man >> plac, nas >> voic

This hierarchy, in turn, implies the following ordering of IDENT(f) constraints for consonants:

(48) IDENT(root) >> IDENT(man) >> IDENT(plac), IDENT(nas) >> IDENT(voic)

I hope this subsection successfully outlined an adaptation of the feature tree in autosegmental phonology as an articulatorily-based model to an acoustic model based on the influence of features on perceived similarity in puns. In sum, the resulting hierarchy is intended as a theoretical basis, strongly related to Lutz and Greene's (2001) sonority-based differences (cf. section 4.6), for a model approximating the actual cost calculated on the basis of frequency (cf. subsection 4.5.3). Obviously, more work is needed to convert the feature tree model into a fully specified and theoretically sound auditory-acoustic variant.

Note that a given pun-target phoneme pair in punning correspondence will normally violate several of such IDENT(f) constraints. This ensures for particularly disruptive such cases, that the cost incurred is sufficient to make the violation worse than that of a generally higher MAX or even higher DEP constraint. This is all the more relevant where additional costs for environmental and suprasegmental constraint violations (cf. subsection 5.3.4 and section 5.4) are factored in. For example, a pair that violates all IDENT(man), IDENT(plac), IDENT(nas), and IDENT(voic) constraint, like θ -> n, is assumed more costly to perceived similarity than the deletion of a sonorant in a consonant cluster. This is, of course, a major deviation from OT, where the violation of a higher ranked constrained renders all violations of lower ranked constraints irrelevant for the evaluation of a given candidate. I will return to this point in more detail in section 5.5. below.

5.3.3 Vowel Space

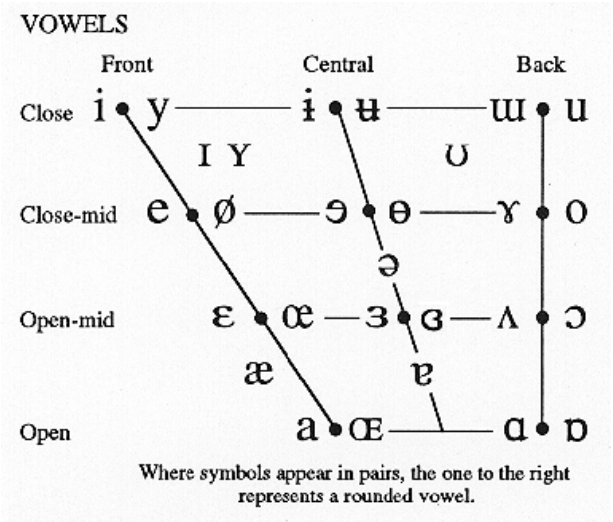
For vowels, a model based on the traditional trapezoidal representation has been found to sufficiently approximate the distribution of contrasts in the corpus. In general, models to distinguish vowel phonemes are more acoustically based than models for consonants, so that a more straightforward implementation for punning sound similarity, which is largely, but not exclusively, acoustically perceived sound similarity, is possible.

Traditionally, vowels were described by charting them along two axes thought to correspond to the position of the tongue, a horizontal one representing how for to the front or back the tongue is situated when the vowel is produced, and a vertical one, representing how high or low the tongue is raised. Charted along these two axes, the vowels were thought to be situated in a ‘space’ that is circumscribed by a trapezoid. With

the invention of the X-ray it became clear that the highest points of the tongue were much more along an axis for the vowels of English (cf. Ladefoged 2001: 115). But the trapezoidal model is still used, but with the underlying theory of frequencies of the first two vowel formants frequencies that are used for its axes. Without going into more detail about the acoustics of formants (for that, see, for example, Ladefoged 2001: 31-46), it should be clear that the current use of the trapezoidal 'vowel space' in phonetics is acoustic in nature (frequencies) and not articulatory (tongue position).

The following Table (6) shows the standard vowel chart of the International Phonetic Association (IPA), reproducing the trapezoidal form, and the metaphoric labeling of the axes as front vs. back (horizontal) and close vs. open (vertical). What it represents is rather the frequency of the second formant of the respective vowel as the horizontal axis (non-linearly from ca. 2200 Hz for [i] via ca. 130 Hz for [a] to ca. 700 Hz for [u]) charted against the first formant on the vertical axis (linearly from ca. 300 Hz for [i] to ca. 800 Hz for [z], cf. Ladefoged 2001: 43).

Table (6): IPA Vowel Chart



In the following chart (49) I have simplified the metaphorical distances of the vowels on this formant chart into values for metaphorical space traversed when one is changed into the other.

(49)	**i/ɪ**	0.15	.	0.15	**u/ʊ**
	0.1	0.1	0.1	0.1	0.1
	e/ɛ	0.1	**ə**	0.1	**o/ɔ**
	0.05	0.05	0.05	0.05	0.05
	æ	0.1	0.1	0.1	0.1
	0.05	0.05	0.05	0.05	0.05
	.	0.1	**a**	0.1	**ʌ**

Accordingly, the cost of the violation of vowel constraint identity is assumed to be equal to the smallest possible sum of (non-bolded) figures one has to traverse to get from the pun vowel to the target vowel. This model is thus based on a straightforward metaphor of locality in the simplified vowel chart. Additional differences incur an additional cost of 0.05, namely, the tense/lax distinction, and the difference for an on-, or offglide in a diphthong as compared to its simple vowel counterpart. Note the higher values along the top part of the horizontal axis that are intended to account for the greater frequency distance of second formants against low frequencies of the first formant.

5.3.4 Environment of Constraint Violations

When the types of non-identical correspondent segment pairs of targets and puns are considered in their immediate intrasyllabic environments,[37] several tendencies can be observed that need to be factored into future extensions of the present study to improve its accuracy in modeling punning sound similarity. To facilitate the evaluation of specific environments in which a given violation appeared to occur overproportionately frequently, it became obvious very soon that it was necessary to establish a baseline proportion of a phoneme's occurrence in certain environments, especially for consonants. A simplifying approach to this was pursued by assuming specific relevance of four environments for perceived sound similarity: onsets and codas for consonants, especially with respect to consonant clusters, and beginnings and ends of words, both for consonants and vowels. The assumption here is that overproportionately frequent changes for a given environment should be made less costly, that is in terms of the underlying constraint hierarchy, the corresponding correspondence constraints should rank lower than those for the same non-correspondence in an environment in which the phoneme affected is proportionately as frequently found as it is violated.

The corpus of violation pair types reproduced in Appendix D was evaluated for overproportionately occurring environments. For consonantal pairs the baseline given in Appendix C was established by importing the CMUDict entries into a database to make them searchable and sortable. In order of prominence, the following significant effects

[37] While the syllable has not been implemented as an automatically tagged level of analysis in the present study, such an extension is a readily available as an add-on (cf. Hammond 1997).

were observed, all of which are in accordance with general phonological mechanics (cf. also section 5.1):

1. Word-initial non-correspondence of consonants, in general, seem to be more tolerable, less affecting perceived similarity in punning than non-correspondence elsewhere.

2. Consonants in onset clusters are more volatile than the same consonants carrying the onset by themselves.

3. Parallel, consonants in coda clusters show the same tendency to allow for violation more readily.

In addition, the following less significant observations could be made from the immediate intrasyllabic and word-boundary environments: Related to 1., insertion (violation of DEP_{PT}) of any segment, regardless of sonorance status, is more frequent at the beginning of words than at the end of words. Insertion or deletion of [j] before the vowel nucleus, especially high back vowels is frequent. Insertion or deletion of [r] in word-final position after a vowel is less disruptive to perceived punning similarity than elsewhere. The schwa alternates with its fully articulated pendants in target-pun pairs overproportionately frequently before syllabic consonants [l], [r], [m], [n].

5.4 Suprasegmental Constraints

Violation of the constraint on identity in the position of lexical stress, IDENT_{PT}-STRESSED-V, is extremely rare, indicating a high significance of stress for perceived similarity in punning, and, accordingly, a high cost is associated with it. In 39 instances the stress is positioned later in the pun than in its target, e.g., *cover -> covert* (marked “l” in the penultimate column in Appendix E), and in 24 instances earlier, e.g., *important -> impotent* (marked “e”).

Similarly, the number of syllables is identical in the overwhelming number of target-pun pairs, i.e., the foot-structure is preserved. Where there is a difference, a significant asymmetry can be observed. In only 23 instances does the pun have one syllable less than the target, e.g., (marked "-1" in the final column in Appendix E), violating a MAX_{PT}-σ constraint. But in 49 cases, the pun has an additional syllable to the target (marked "+1"), of which two even show two additional ones ("+2"). The predominance of this latter type, violating the DEP_{PT}-σ constraint, e.g., *rickshaw* -> *ricochet*, seems to indicate the influence of recoverability, as it must be assumed that to identify the part of a pun that is repeated in the target is less costly than to infer additional material in the target that is not available in the pun.

In sum, the constraints on stress and syllable identity must be assumed to be violable only under a high penalty in terms of punning similarity. Accordingly, these constraints are ranking high, with the following internal hierarchy:[38]

(50) Max_{PT}-σ >> Dep_{PT}-σ >> Ident_{PT}-stressed-V

5.5 Cumulative Cost of Constraint Violation

Another central assumption is that a longer punning domain is more robust than a shorter one. It is intuitively clear that a target-pun pair consisting of three sounds cannot function if all three sounds are different, whereas a pair consisting of 14 sound segments very well may tolerate violation of root identity of three of its sounds. Lutz and Greene

[38] On the crucial assumption that relative frequency in the corpus is equal to lower ranking in the punning correspondence constraint hierarchy, the main difference of the present study to canonical OT, see subsection 4.5.3 and section 4.6.

(2001: n.p.) also have identified this as a major problem preventing the computational implementation of OT:

> As such, in OT constraint violations are infinite. If constraint C1 dominates constraint C2, no number of violations of C2 can ever overcome a single violation of C1. The power of the mechanism of constraint domination has been well argued. Its framework and general results have proven useful here as well, but in the end, for the purpose of measuring phonological similarity, constraint violation must be quantified in some way as something other than infinite. It is in this sense that the evaluation procedure most significantly departs from the principles of OT.

In sum, the combined violation of several constraints must be dealt with differently here than in standard OT. There a hierarchy is sufficient, in which the violation of a single high constraint ultimately decides on the most optimal candidate for a given input form. In OO constraints evaluating the similarity of several output candidates, which have already survived a run of IO evaluation witnessed by the fact that they are surface English words, a cumulative penalty must be assigned. This becomes most clear, when one considers that the multiple violation of the same constraint in the same candidate must carry a higher penalty than its single violation, and possibly a higher penalty than the violation of a more important constraint.

To account for this effect, the total cost incurred by the violations in a pair has to be put in relation to this length factor. In other words, longer target words, providing more of a torso that supports recoverability, allow for puns committing more violations. A ratio of violations per segment must be factored in. At his stage, the total is divided by a quotient linearly dependent on the sound segment number of the target multiplied with an arbitrary factor to render the result more intuitively understandable by keeping it between 0 and 1.

The next subsection will summarize these assumptions into a formula, which will necessarily oversimplify the complex issue of constraint interaction. But from the previous discussion in this chapter it will be clear that a full enumeration of punning output-output constraints is an impossible undertaking, as well as an unnecessary one, as such a model would not be implementable. The number of constraints themselves is, theoretically as large as the number of phoneme contrasts possible, namely 25^2 for the consonants, and 18^2 for the vowels, multiplied with the relevant contexts, in addition to the suprasegmental constraints. Obviously, the relation of the model to tendencies among groups of constraints must remain the level of detail meaningful for the present discussion.

5.6 Evaluation of Punning Output-Output Correspondence

Based on the model described in the previous sections, each individual **cost value** *C* for non-correspondence of a pun with respect to its target is computed in terms

- of the **IDENT constraint violation**(s) *v* of featural branches of the autosegmental feature tree, already a compositional factor varying according to
    - the acoustic feature tree branches involved, including
    - the direction of the violation (P->T, P->T), and where it occurs, where the
    - **DEP and MAX violations** are conceptualized as special particularly costly instances of IDENT violation, to accommodate them into the same table,
- multiplied with the **environmental factor** *e* in terms of
    - syllable position,
    - neighboring segments,
- and added to **suprasegmental factors**, namely
    - the DEP/MAX-□ violation □ for each string, if any, and
    - the IDENT(‘) violation □ for each string, if any
- divided by the **punning domain length** of the target in terms of number of segments *s* and syllabic structure multiplied by a factor *x*

resulting in the following formula:

(51)
$$\frac{(v_1 \text{ x } e_1) + (v_2 \text{ x } e_2) + \ldots + (v_n \text{ x } e_n) + \square + \square}{xs} = C$$

The following section will outline the implementation and improvement of the assumptions on which this formula is based.

5.7 Implemented Edit Distance Computation

The actual formalization of output-output correspondence in punning which is to be implemented can't be constraint-based as this is prohibitively unwieldy as a computational model. The main reason is that for the ranking of more than one output candidate, an accumulation of the costs incurred through constraint violations is necessary. Therefore, a cost table, that represents the constraint hierarchy will serve as the interface of the theoretical model and a program that computes the edit distance between the target and pun.[39] The cost table itself is documented in Appendix G. In its current form it is based primarily on the frequencies calculated in Sobkowiak (1991) and repeated here in Appendix F. These frequencies are translated into costs under the assumption that the more frequently a non-corresponding phoneme pair occurs, the less negative effect it has on the perceived similarity between target and pun. The initial specific algorithm to translate frequencies into cost are also given in Appendix G. The different costs are implied by the ranking of constraints in a hierarchy in relation to features applied to phoneme pairs in context. Thus, the accumulation of cost utilizes the ranking and, in turn, implies it. In view of this, the main aim of this section is to describe

[39] I am extremely grateful to Krista Bennett for writing this edit distance program and suggesting many improvements without which this part of the present study would have not been possible.

the computation of the edit distance (Hall and Dowling 1980) and the components required.

The main tools for the computation of the distance between the target and pun in a given pair as performed on a Macintosh G3 computer both in OS X and its underlying UNIX system using the command line interface are:

1. the database of 1182 puns, processed in ACI 4th Dimension, from which the unique number and the transcription of each target and pun are exported into a text file;
2. the cost table based on observed frequencies, processed in Microsoft Excel, and exported into a text file;
3. the program to compute the edit distance by looking up the cost for each pair, written in Java, automatically producing two types of output,
    a. the sum of the cost values for each target-pun pair, and
    b. the individual violations and their cost, and the segmental context in which it occurs.

In addition, the exported text filed to be purged of OS specific formatting characters, for example <^M>, before they could be fed to the Java program, using a simple script.[40]

As mentioned, the cost table was created using a table that had frequencies (x) of the violations and assigned the cost (y) according to the following formulae:

for the consonants, asymptotically, namely: $y=x^{-0.6}$; for vowels, linearly, namely: 0.3-0.3x/161, 161 being the most frequent violation and 0.3 the highest violation cost vowels should carry in view of their assumed influence on punning similarity. These formulae are initial hypotheses, which will have to be adjusted in future implementations, which will be much facilitated by the fact that any input file with pun-target pairs can now automatically be processed with the existing edit distance program.

[40] I am grateful to Andreas Waschbüsch for adapting a script that took care of this.

Examples for the two types of output provided by the program are given below,[41] (52) showing the sum of the costs and (53) showing the individual costs for contrasts:

(52)	number	target	pun	total cost	adjusted cost	contrasted pairs	
	9	spjuri0s	sp.8ri0s	0.38	0.19	j-.	u-8

(53)	number	target	pun	contrast	environment	cost
	9	spjuri0s	sp.8ri0s	j-.	p_u	0.2
	9	spjuri0s	sp.8ri0s	u-8	j_r	0.18

These different outputs are then re-imported into different tables of the relational database for integration with the constraints that still have to be hand-tagged, and interpretation towards the reevaluation of the cost table and the model on which it is built.

5.8 Overcoming the Infinity Problem of OT

A final note is required, before we can turn to extra-phonological factors of imperfect punning in the next chapter. The purpose of the present study is to develop the model for punning similarity on the basis of linguistic theory to make it available for actual implementation in natural language processing.[42] While it is itself not a part of the research presented here, the computational implementation makes it necessary that this research must produce a result at a level of formalization high and complete enough for a consumer as dumb as a computer to be able to digest it. In this respect, it is important to highlight a crucial difference between input-output constraints in canonical OT and output-output constraints like those that hold between an imperfect pun and its target. Canonical OT, unlike other generative theories in linguistics, suffers from the crippling

[41] The target and pun are given in the ASCII transcription of the Sobkowiak corpus (see also section 6.1), where a period stands for a segment missing in the correspondent.
[42] See Albro (2003) and Tesar (1995) for a more canonical approaches to OT-based computational modeling.

assumption that its constraint hierarchy evaluates an infinite number of candidates generated from the input, referred to as the “Richness of the Base” (cf. Archangeli 1996: 14, 27). This makes OT in principle impossible to implement. If, on the other hand, an evaluation hierarchy is built to evaluate a finite set against one form, as in my approach the set of existing English words selected at a meaningful level of linguistic abstraction and with respect to the appropriate linguistic parameters, it can become a useful tool. In a future implementation it should be attempted to restrict the candidate set of puns evaluated against the input target even more, by using only words that denote concepts that are in humorous relationship to the concept denoted by the target, that is, make the target-pun pair not just phonologically but also semantically well-formed in the sense described in chapter 2.

6. RECOVERABILITY BEYOND PHONOLOGY

6.1 Introduction: Interaction of Phonological, Syntactic, and Semantic Factors

The previous chapters have outlined the relatively more developed research that has produced results for many of the phonological factors involved in heterophonic punning towards a more refined model based on new theoretical assumptions, both phonological and humor theoretical. This chapter will venture on the remaining linguistic factors deemed relevant in imperfect punning, in particular at the syntactic level, namely word classes, grammaticality, and formulaicity, as well as the interaction of these factors with a view to recoverability of the target of puns, towards the integration of all relevant target recovery factors into one unified framework. As argued above, recoverability is strongly covariant to, but not identical to phonological similarity, which is only *one* factor that facilitates it. To repeat, a pun must be both, semantically well-formed, having two opposed, overlapping scripts, as well as recoverable, i.e., the target must be relatable to the pun with the help of factors best analyzed separately at the levels of semantics, syntax, morphology, and phonology.

In this section it will become particularly obvious that I am not using a model inspired by Optimality Theory (OT) for the sake of OT, but in order to provide a meaningful account of paronomasic punning. Obviously, such an account must include more than just phonological factors. And if not included as part of a full analysis, then those factors, as far as they have been identified, will be mentioned again and their role

outlined. To do this, i.e., to give the necessary breadth to the analysis of a broad phenomenon without getting lost in the problem of infinite levels of analysis, some of the depth that phonological analyses can normally afford to devote to the full exploration of their tools has to be sacrificed for the sake of the useful exploration of the puns themselves. Accordingly, it is my explicit assumption that this is not so much a sacrifice of my discussion overall, but rather a gain. And what I claim to gain is real insight into the complexity of punning and the interaction of the factors at work, even if I have to be eclectic about the use of theories I employ, fitting several connected engines under the hood of a theoretical vehicle explicitly aimed at application. In this sense, it is assumed possible that OT can be replaced in my framework by a different, better theory, should one be identified or become available, without losing the ground gained by my approach. I believe that humor is primarily a semantic issue, so a future extension to semantic priming and recoverability of phonological dissimilarity based on the discussion above is expected to model the issue of punning with respect to these levels of analysis.

The main goal to be pursued in this chapter is based on the hypothesis that the recoverability of the target of a pun must be modeled as a compositional value. The assumption is that if there is little sound similarity, then there must be more syntactic or semantic clues that enable the hearer to infer the target. The following example may, again, not be a good joke, but it is a good example for little phonological similarity between pun and target, which is nevertheless salvaged with the help of a collocation:

(54) Following the Watergate incident in Washington, there were so many lawyers trying to get involved in the settlement of conflicting evidence that they became

known as "ambivalence chasers." (12)
ambivalence -> ambulance, &mb**I**v0l0ns -> &mb.**j**0l0ns[43]

Between *ambivalence* and *ambulance*, there is not only the contrast of phonemic segments typical for heterophonic punning (highlighted by boldface), but through the lack of the nucleus of the second syllable, and with it the whole syllable, there is also a different position of main stress in *ambivalence* (highlighted by underlining). Accordingly, my hypothesis is that this large sound dissimilarity should block the recoverability of the target *ambulance*—having a different stress pattern and one syllable less than its pun—, but that it is crucially counterbalanced by the cohesive force of the collocation *ambulance chaser* triggering its first part *ambulance*. That is, the recovery force of the formulaic language, in this case the collocation "ambulance chaser," takes over much of the responsibility that would otherwise have to be carried by phonological similarity alone. The force of the formula is, of course, working together with the semantic priming of the script LYING that is triggered by *Watergate* and includes *ambivalence*.

In this chapter I will describe these non-phonological recovery factors as they have been encountered in the full tagging of my corpus of roughly 1,200 puns. Because the aim is to devise a fully formalized model towards automatic generation, more frequent and therefore productive mechanisms will receive more attention and rare ones will only be briefly mentioned, but not systematically integrated into the present approach.

---

[43] In this chapter many complete examples from Crosbie (1977) and elsewhere will be quoted. The notational conventions in these examples are the following: After the full punning joke text, the pun and target are given in italics; if the sound similarity is relevant for the point of the example, their transcription is provided either in square brackets in standard IPA, or without brackets in the ASCII IPA adopted by Sobkowiak (1991) if the example is from his corpus. If the citation in round brackets consists only of a single number, it refers to page number in Crosbie (1977), otherwise the standard citation is provided.

6.2 Quasi-punning non-phonological Recoverability

It has to be emphasized again that not any distance between two forms is acceptable in punning, even if syntactic and semantic support makes the target recoverable. As discussed in detail in section 2.2, without at least a token amount of sound similarity, the logical mechanism (LM) of cratylism would be lost. In the following example, we have no problem in recovering the intended target of *rosé*. Yet, the example is not a pun, but a different, if similar, form of wordplay. That is, while sharing the support of syntactic and semantic context towards the recoverability of a form—here *blue*—and all formal requirements for a joke—namely an SO of high vs. low scripts along the lines of PATRIOTISM vs. ALCOHOL—it is not within the purview of the present study, because the LM here is not based on cratylism. It works similar to straightforward juxtaposition LMs (cf. Attardo, Hempelmann, and Di Maio 2002), only that the juxtaposition is paradigmatic here:

(55) America: Land of the Red, White and Rosé
Rosé -> Blue (Time 2003: 83; headline of a short article on wines produced in the United States)

The inverse consequence of the salvaging effect is that there can be purely phonological pseudopuns, mostly with homophony or very little heterophony, which completely lack syntactic embedding, or, worse, even lack a second script. These are not only often hard to get for a non-native speaker, but also often not funny, or even not jokes at all, because they lack a second script, or have no script opposition, which is among the requirements of Raskin (1985). Many non-joke puns are found in the "definition" category (see sections 6.10 below and 2.3 above). There are approximate cases to the non-joke constellation that nicely illustrate the interaction of recoverability factors. The

following word-play can be found in Crosbie's collection. It is not a pun, because there is no opposition of overlapping scripts, i.e., no semantic incongruity, only a minor phonological one:

(56) As the weatherman said, "Disgusting is going to come to blows." (67)
disgusting -> *this gusting*, dIsg^stIN -> DIsg^stIN

In this quasi-joke there is extremely little phonological distance between the target and the pun, in only two features of the onset of the first syllable: place slightly shifts from alveolar to interdental and a tad more continuous quality is added in the target.[44] But, more importantly, there is no semantic joke quality, no incongruity between the meanings of *disgusting* and *this gusting*, but mere revelry in wordplay through the relation of *gust* and *blow*, in the idiomatic phrase *come to blows*. While these factors support the recoverability of the target, they don't make this text a joke.

In contrast to this example of word-play, I will discuss the interacting factors supporting the target recoverability in punning joke texts in the following sections of this chapter, starting with morphological issues in the punning domain.

6.3 Intermediate Level: Morphology

In the present study I will devote only one short section to morphology. This has two very different reasons, in addition to the low frequency of puns that involve morphology in the corpus. The first is that to a large degree, morphology is a function of syntax as it is understood here (and, for different reasons, in current syntactic research).

[44] The little distance that the feature-based analysis provides shows again how much articulatory bias the feature system has, when the pun and target are indeed homophonous in some dialects.

The second reason is that morphology is a very regular subsystem of the English language and because of this regularity has been sufficiently formalized as a NLP task (see, for example, Karttunen 1983, Dalrymple et al. 1987, Ritchie et al. 1992).

At the level of pun generation that my approach ultimately aims at, a morphology module that can increase and enhance the output would simply add the possible inflectional suffixes to a potential pun in order to increase the candidate set with similarity to the target, when the one has a final sequence of sounds that is similar to the other plus the inflectional morpheme.

(57) Once the cooking problem was ironed out, the little fish smelt good. (213)
smelt -> smelled, *smell* + *ed* (past)

In this example, we see not only one of the numerous names of more or less obscure fish that Crosbie (1977) used in many of his puns (*smelt*=small salmonoid fish of the *osmeridae* family), but a pun joke the tense of which is simple past, because the pun *smelt -> smelled* (*smell* + past -*ed*) is more similar than that on *smelt -> smells* or *smelt -> smell*. The latter two would possible candidates were the joke text in the simple present.

An additional evolution of the evaluation model, and the envisaged joke generator based on it (see chapter 7), could include derivational affixes, but their distribution is far more complex—namely meaning-related—than that of inflectional morphemes.

(58) The lobe was the original earring aid. (74)
earring -> hearing

In order to match *earring* and *hear* in this example, the latter would have to be turned into the deverbal noun *hearing* through the derivational morpheme -*ing*.

The following is an example from the twelve derivational morpheme puns that make up a mere 1% of the corpus:

(59) An optimist is an anti-sceptic. (212)
(anti)*sceptic* ->(anti)*septic*

As mentioned above, because of their low productivity the expenditure to model these cases fully does not seem warranted.

In sum, these morpheme puns illustrate at short and hence very forceful range an important factor for the recoverability of heterophonic puns that cannot be ignored in any adequate description. Namely, that the pun is often embedded into a larger structure, that statistically significantly often only has the target in the syntagmatic place in general, where the specific joke text has the pun. In other words, the pun and target are in a paradigmatic relationship that crucially enables the recovery of the target. More specifically, at the morphological level, the prefix *anti*, for example, can combine with most adjectives and nouns, of these combinations only 129 became so fossilized that separate entries are devoted to them in a standard dictionary (American Heritage College, 3rd edition: from *antiaircraft* to *antiwhite*), indicating that these have higher potential to be recognized as targets than those that haven't become fixed forms.

6.4 Superlexemic, Multiple-Word Domain Punning

The corpus contains only few puns involving superlexemic semantic contrast at the morpheme level, and they are the rare exception among almost exclusively word-level puns and few multi-word puns. Of the latter—jokes in which the contrasted segment crosses word boundaries in the pun, the target, or both—only one combination is of numerical relevance, namely those where a single-word pun is mapped onto a multi-word target. Of such pun-target pairs 216 are in the corpus (18.02%), most of which are two-

word targets (173), many three-word (40), few four-word (3) and one even five-word (*onomatopoeia -> I don't want to pee*; 165). There are only 30 multi-word puns that are mapped onto single-word targets and 16 multiword puns that are mapped onto multiword targets. Interestingly, of the latter only one maintains the word boundaries from pun to target (*knock need -> knock knee*; 158); in the others, usually a coda of a word in the pun becomes an onset of the subsequent word in the target (e.g., *an oaf -> a loaf*; 163) or, inversely, pun onsets become target codas (e.g., *wore cologne -> walk alone*; 52).

An easily applied and hence frequent mechanism (40 instances) that involves crossing of word boundaries are pun-target pairs of nouns, one of which has an additional article, while the other's first syllable is phonologically similar to *the, a* or *an*:

(60) A publisher who specializes in novels about zombies, goblins and demons has just established the Dybbuk-of-the-Month Club. (72)
Dybbuk -> the Book

This mechanism should be included in a pun generator because of its frequency and easy implementation, once the phonological component has been assembled.

After these introductory sections providing an overview over some rarer mechanisms, let us now turn to a more general discussion of the superlexemic, syntactic framework in which the non-phonological recovery factors interact, before we continue to look a the syntactic recovery support in punning in more detail.

6.5 The Context as Torso

Hausmann calls the incomplete syntagmatic corset that the text a punning joke provides a "torso" (1974: 22f) that provides sufficient co-text to deduce missing parts and make them paradigmatically available through the obvious gap or ungrammatical link in

the syntagmatic chain. This general observation and its obvious importance for making sense of language—predicting new input on the basis of already processed input—has also been mentioned by Preisendanz in the connection of verbal humor: "in the medium of language, that what is not uttered brings about the virtual presence of the missing complements" (1970: 23; my translation).

Hausmann illustrates the concept of vertical realization (paradigmatic availability) of what he calls "quasi-homonyms," i.e., heteronyms/-phones, with the following example:

(61) Un parti révolutionnaire, c'est un parti d'extreme ... (1974: 22)
(A revolutionary party, that's a party of the extreme ...)

The force of the idiom *parti d'extreme gauche* (party of the extreme left) within the semantic field of "political party," in particular, "left-wing party," as triggered by *revolutionary*, reduces the possibility of what could appear in the gap at the end of the phrase not only syntactically to a noun, but, crucially, also semantically. The syntactic restrictions pertaining in idiomatic pun jokes in specific and pun jokes in general will be discussed in more detail in a larger subsection below.

The recovery of the missing word in example (31) would receive additional support for its recoverability if a similar sounding form were inserted in the position where *gauche* is expected. Hausmann's example is the nonce form *dauche*. The following version of the example has all features of a heterophonic pun joke, except for the most important one. As there is no meaning in *dauche*, there is consequently no second semantic script that could be in opposition and overlap with another script as triggered by the remaining text with the recovered target *gauche* inserted:

(62) Un parti révolutionnaire, c'est un parti d'extreme dauche (1974: 24)

But the next version of the example includes yet another force at work also in the example as it was indeed found, not as an intended pun, but with a typo that happens to be semantically felicitous as a joke and as such used embedded into another joke:

(63) Un parti révolutionnaire, c'est un parti d'extreme fauche (1974: 25)

The pun *fauche* ("petty theft"), which is the form actually found in Hausmann's corpus where it is presented as a typo, shows the required semantic overlap, although Hausmann seems to misinterpret it as *fauché* ("penniless"), when he analyzes the overlap to lie in "an undeniable solidarity between poverty and revolutionary spirit" (1974: 25)[45]. But the additional force I will focus on in a separate section below is the ungrammaticality, in particular on syntactic grounds.

In general, the recovery of the target in heterophonic puns is just a specific case of the complex task of hearing (and, with some differences, that of reading in the case of eye puns) that we constantly master successfully. Each time a new segment is to be understood, it has already been cued, most importantly semantically, partially as encoded by and in addition to the syntax of the preceding text. We expect a member of a class of words belonging to a specific semantic script and of a specific syntactic word class, maybe even a specific lexeme including its specific morphological realization as appropriate for the syntactic position. And if anything remotely close to it is produced, we perceive not just the pun, but also the expected target.

In heterophonic jokes we have the additional effect that not only the actually uttered segment is perceived, but because it differs minutely, yet crucially, from the expected

[45] I am grateful to Shaun Hughes who made me aware of this mistranslation by Hausmann. Accordingly, the joke is not so much expressing revolutionary sentiment—the Communist Party as the savior of the penniless—but an anti-Communist stereotype—the Communist Party as the thief destroying the wealth created by the capitalist system.

segment, the latter is also recovered. Understood this way, heterophonic punning involves an extremely easy variant of a CLOZE readability test (cf. Taylor 1953) or, if we understand the LM as an enabler of abductive leaping from one script to the other, the target recovery as facilitated by the torso can be understood as "overcoded abduction" (cf. Eco 1983).

But jokes, as a specific text type, possess an additional level of semantic cueing: Since the hearer knows the text is a joke, if the appropriate non-bona-fide mode of communication has been established, he or she is expecting opposite overlapping scripts to be triggered by the joke. And heterophonic pun texts involve a specific level of joke expectation, namely, that one segment of the text, usually towards the end, will not only trigger the second of these scripts, but also that it will do so by virtue of being similar in sound to another segment that will make the joke text semantically well-formed as a text in general and a joke-text in particular, as well as syntactically well-formed. In short, it has to be grammatical at all levels of language, and, additionally, fulfill the criteria for being a joke discussed in detail in chapter 2.

6.6 Idiomatic Expressions: Semantic-Syntactic Interface

This section focuses on the role of such formulaic idiomatic patterns as they abound in the corpus and prove that beyond sound similarity and semantic script affiliation, syntactic-semantic formulae are of high relevance to the recovery of the targets in heterophonic punning (cf. Sobkowiak 1991: 118, Zwicky and Zwicky 1986: 500). The main hypothesis here is that "[n]ot only do formulaic puns as a whole show more paronymy; the also have wilder paronymy" (Lagerquist 1980: 189), because they can.

A total of 798 pun jokes involve one kind or another of formulaic torso into which the pun is embedded, and in view of the likely unfamiliarity with some formulae of English and the American culture on the part of this author, this figure can only have to be adjusted upwards. From the 1182 heterophonic puns under investigation here, at least 67% involve formulaic patterns. Lagerquist finds a similar significance of formulaicity for paronomasia in his corpus (Cerf 1968), where he calculates that formulaic pun texts are more susceptible to phonetic distance, i.e., heterophony (64%) than other pun texts where only 45% are paronomasic (1980: 189).

Formulae include phrasal verbs, collocations, idioms, proverbs, quotes/titles, and clichés, in short, "any grammatical form [for which] the meaning is not deductible from its structure" (Hockett 1958: 172). Formulaic language is the clearest case for the non-compositionality of semantics, because the meaning of the whole phrase can't be understood as the sum of the meanings of its parts. At one end of the spectrum we find purely syntactic fixation, for example, an object missing for a transitive verb, combined with a semantic specification about the kind of object required, for example, only edibles can the object of *eat*. On the other end, we have a quote, each single element of which is fixed and fully recoverable, like the much-abused "the rain in Spain falls mainly on the plain" from *My Fair Lady*, featured in five of Crosbie's texts (I hesitate, again, to call the following example of wordplay a joke). Note the heteronymy across four segments that a quote formula can recover:

(64) If you want to do something about altering your cat, it may be consoling to know that the pain in spaying is mostly in the brain. (216; see also two more examples on that page and 57, 249)
the pain in spaying is mostly in the brain -> the rain in Spain falls mostly on the

plain;
pain -> rain, *spaying -> Spain*, *in -> on*, *brain -> plain*

Within this spectrum, two main types of idiom have traditionally been defined as illustrated by Bolinger's (1971: 111-31) distinction of first- and second-level stereotyping (in that particular case, of phrasal verbs):

(65) first level: can (at least partially) be explained from its components
second level: completely opaque, non-compositional meaning

This distinction is obviously too broad. To account for the recovery force of different degrees of fossilization in idiomatic speech, I chose a subset of Makkai's (1972) more finely graded hierarchy. For further taxonomies with mixed criteria see the synopsis in Hausmann (1974: 51-54) and a more recent approach in Moon (1998).

In general, Makkai also distinguishes two areas of idioms, the lexemic and the sememic (1974: 117ff), corresponding to the traditional distinction given in (65). Examples of the first type are phrasal verbs like *fill up, take down, buzz off*, examples of the second, proverbs like *never steal anything small*. But this distinction is not utilized in my analysis. The reason lies in the centrality of what Makkai calls the "sememic force" that permeates the spectrum of idioms that I am focusing on: The lexemic force enables the recovery of the target when the pun is embedded into an idiom such that its idiomaticity is violated. Such unidiomatic cases call for repair through replacement of the wrongly inserted pun through the paradigmatically available target, forced through the incompleteness of the idiom's torso in the pun joke.

Makkai explains why his distinction between merely phraseologically fixed formulae and genuine semantic idioms for which "synchronic [and often also diachronic] componentialization fails" (1972: 50) will turn out to be secondary for my discussion of

the recovering force of idioms: All "idioms of decoding" (semantic, truly non-compositional ones) are at the same time necessarily "idioms of encoding" (fixed strings of lexemes), actually a subset of the latter that is even more resistant to the rearrangement and transformation. Accordingly, the very force relevant for the recovery of a pun target is inherent to both types.

Of the criteria to categorize types of formulaic expressions, like inflexibility, conventionality, or non-compositionality/idiomaticity, only the inflexibility in syntactic and/or lexematic terms will be understood to incorporate—and, in turn, be affected by—the semantic non-compositionality. Conceptualized in this way, it is most relevant for the recoverability support I am interested in, because my application will allow for more phonological dissimilarity for a pun made more recoverable as a part of an idiomatic torso. For the full implementation of an idiomaticity hierarchy a lexicon of idioms would be necessary against which the input target could be checked. If a pun from within an idiom can be found, the C constraint, that is the cut-off point for unusable targets can be relaxed, that is moved lower (respectively a negative value added to the dissimilarity cost computed from the evaluation formula outline in section 5.6), so as to include that target despite larger dissimilarity on the phonological level (and also the humorous semantic level which is not necessarily interacting with the semantic non-compositionality/metaphoricity of the idiom).

Starting from Makkai's categories, I propose the following provisional hierarchy of idiom-types with lower flexibility/longer sequence/higher recoverability at the top and the number of instances in the corpus identified to fall into them on the right:

Table (7): Types and Tokens of Formulae

1.	quotes	145
2.	proverbs	82
3.	clichés	146
4.	non-compositional, opaque, 2nd area idioms	135
5.	compositional, transparent, 1st area idioms	290
	total	798

Quotes are completely fixed and, consequently, are assumed to have the highest recoverability force. The highest possible phonological or semantic pun-target distance should be possible, but obviously not necessary, here:

(66) The British rain song: A Froggy Day in London Town. (93)
froggy ->foggy; a foggy day in London town (song by Gershwin)

(67) An afternoon snack: the pause that refleshes. (89)
refleshes -> refreshes; "the pause that refreshes" (a 1947 Coca Cola slogan)[46]

Proverbs are almost fully fixed complete utterances with largely non-compositional meaning, namely a moral, usually metaphorical in nature, i.e., pertaining to a different domain than the literal utterance. The only variability allowed here is that of adjusting minor categories, like tense, number, or gender, resulting in minor morphological difference from the prototype:

(68) Asked why he preferred the more mature hens in the barnyard, the rooster explained, "There's no fowl like an old fowl!" (92; one of Crosbie's favorites, although he was only 57 at the time of publication; 97, 167, 179, 233, 261)
fowl -> fool; "there's no fool like an old fool"

(69) Unctions speak louder than words. (240)
unctions -> actions; "actions speak louder than words"

Clichés are less fixed longer phrases (more than three lexemes), but not full sentences, with a non-compositional meaning:

[46] In this section, the formula used in a pun example is given in quotation marks, including its source, if it could be identified.

(70) The song of the fisherman's nightmare: "To Dream the Impossible Bream." (34; another fish pun)
bream -> dream; "dream the impossible dream"

(71) She thought her new boyfriend was creative but when he finally got down to it he wasn't a Rachmaninov to satisfy her. (190)
a Rachmaninov -> man enough; "not be man enough"

The following is a brief list of distinct types of formulae grouped into the more heterogeneous classes 4 and 5 with examples from the corpus. These classes are not strictly mutually exclusive, which is the main reason why I combined them in my tentative hierarchy in Table (6) above:

(72) **phrasal verb** verb particle combination
rip off (196), *catch up* (130)

tournure more than three lexemes, template: "to X the/a Y" (see Makkai 1974: 148f)
kick the habit (190), *have the heaves* (74)

binomial two coordinated minimal phrases
cause and effect (55), *fish and chips* (87)

collocation premodifier and noun combinations
dying breed (247), *assorted flavors* (215)

compounds noun phrase compounds
public-opinion poll (184), *town cryer* (236)

It should be noted that phrasal verbs—despite their prominent role in the English language—are fairly rare in the corpus in positions interacting with the pun-target domain. The simple reason seems to be that the particles are too short to allow for much variation other than in contexts where they are mapped onto the target or from the pun with the help of additional lexical items. In short, the range of phrasal verbs is too short, and only seven of them have been found in the corpus, four of which use the particle in the punning domain:

(73)	rip off:	rib-off -> ripoff (196)
	pass on:	Parthenon -> passing on (170)
	be on:	yawn -> you're on (259)
	urge on:	erg -> urge (78)
		urchin -> urgin' (240)
	catch up:	kitchen -> catchin' (130)
	wake up:	highway cop -> I wake up (111)

In sum, the higher categories in table (6) a have higher rigidity. If a pun-target segment is embedded into a formula belonging to one of those categories, higher recoverability of the target is assumed, all other factors being equal. As an extreme example, in the sample corpus whole stanzas of songs have been encountered with several pun-target pairs embedded into them. But to conclude this subsection, a somber note is necessary: Until either a lexicon of English idioms labeled for rigidity levels, possibly the ones proposed in this sketch, is set up or a useful tool for the analysis of idiomaticity, short of a full analysis of the whole text into which it is embedded, is created, idiomaticity remains quite literally an academic issue. And that is regretful because of its immense role in heterophonic punning as witnessed by the sample corpus, and language in general.

6.7 Word Classes and Syntactic Functions: Lexico-Semantic Interface

This section will discuss the interaction of word class—or classes—of the pun and target with their syntactic function and with the respective syntactic grammaticality of the pun and target in their shared context. The following interaction between the two parameters are possible and documented in the corpus:

Table (8): Word Classes and Syntactic Functions of Pun and Target

word classes of pun and target	*syntactic functions of pun and target*	#
1. same	same	**748**
2. different	different	**216**
3. different	same	**203**
4. same	different	**0**

The figures in the last column give the number of instances of the combinations out of the 1182 puns in the sample corpus. The three classes actually represented are illustrated by the following examples:

(74) Actually, Arab women really fear the wolf in sheik's clothing. (209)
sheik's (noun, possessive premodifier) -> *sheep's* (noun, possessive premodifier)

(75) As Ruth said to Naomi, "Lesbian ardent pair." (136)
lesbian (noun, possessive premodifier)-> *let's be an* (complex verb, article)

(76) And then there was the traveler who went around the world in a tea daze. (63)
a tea daze (prepositional phrase: article, nominal premodifier, noun, object of the preposition)-> *eighty days* (prepositional phrase: adjective, noun, object of the preposition)

In general, target and pun in punning jokes have the tendency to be in the same word class as well as to fulfill the same syntactic function. A pun generator can maximize its output by focusing on this straightforward kind of pun:

(77) Cleopatra instituted the practice of keeping a handkerchief tucked in her cleavage and thus became famous for taking a wiper to her bosom. (254)
wiper (noun, direct object) -> *viper* (noun, direct object)[47]

But not all pun-target pairs that fulfill the same grammatical function also stay within the same word class. Accordingly, the 1182 pun target pairs were tagged for parts of speech with the CLAWS tag set (see Appendix A) after initially only having been tagged

[47] Another possible, but less accessible, interpretation for the target in this joke, which Shaun Hughes made me aware of, is [waypə], an acceptable pronunciation of *viper* in certain upper-class British dialects. While it is a rare dialect (cf. Trudgill 1990: 45), it remains alive through fictional characters such as Dickens' Sam Weller.

for grammatical function (see below, Appendix B) in order to identify regularities that should be formalized for a pun generation system.

The aim of the tagging of the corpus for word class/part of speech and grammatical function is to identify the common types of crossing of word class from target to pun. This time-consuming procedure was warranted by the fact that word-class would be the only syntactic information available for the input of the type of pun generator. In order to produce meaningful output, the pun and target must be of such word classes that, if they're not identical for both, a future implementation towards a full joke text can still utilize the output of my module.

Word class difference between target and pun is frequent. Of the general type without the special cases discussed in the following subsection, there are 434 target-pun pairs (36.72%; lines 1-6, 10-11 of the following table (8)) that do not show word class identity in the sample corpus. Two different scenarios with respect to grammaticality are found here (see also below). The more common type has grammaticality only for the sentence containing the target, but not the pun (222; 18.78%; lines 1, 4, 7). This grammaticality imbalance, illustrated by the following example, will receive more attention below (section 6.8). In the following example, the clause "hence lay the eggs" is not grammatical, but with the target "hens" in place of its target "hence" it is.

(78) The roosters do the crowing, but the hence lay the eggs. (110)
hence (conjunction) -> *hens* (noun)

The generation of such puns is a necessary feature of a pun generator as they are very frequent. And they have an the additional recovery force that lies in the very ungrammaticality of the pun text and may be required to salvage higher phonological (or

other) dissimilarity. But these puns are considered worse in terms of funniness, because this very “brute force” recovery is considered unsophisticated and many of them, like example (78), have to be considered word play rather than jokes as they lack the necessary semantic component (cf. section 2.3).

Table (9): Synopsis of Word Class (WC), Syntactic Function (SF), Grammaticality (G): — = no, X = yes, / = unclear, B = both, T = only target, P = only pun, ne = not embedded

	WC	**SF**	**G**	#	∑
1	—	—	T	169	
2	—	—	ne	37	
3	—	—	B	10	
					216
4	—	X	T	31	
5	—	X	ne	14	
6	—	X	B	158	
					203
7	X	X	T	22	
8	X	X	ne	74	
9	X	X	B	652	
					748
10	—	—	P	1	
11	—	/	/	14	
					15
					1182

This section focuses on the more elegant and less common, yet still frequent type pun that shows different word class between pun and target, but preserves grammaticality for both by maintaining the syntactic role of both. The corpus contains 162 such punning texts, comprising 13.71%. In order not to lose one eighth of the possible output of the generator it seems warranted to also incorporate such puns in the generation scheme, in the sense that a possible pun will not fall below the C-constraint if it lies in a word class that can take on the same grammatical function that the target has.

An initial plausible hypothesis for this class is that the most frequent such pun-target domains should include parts of speech that can premodify and those that can postmodify the head of a noun phrase, as well as anything that has the same syntactic freedom as an adverbial. What makes the analysis simple and the generation hard is that there are very few purely optional, syntactic choices that affect meaning very little, or, ideally, not at all. One example is the positioning of adverbials that is comparatively free in English.[48] Oaks operates under a similar hypothesis for one of these functions when he mentions that he groups appositives as postmodifiers and non-head nouns in a compound as premodifiers, not distinguishing them further as parts of compound (1990: 9).

Appendix B lists the syntactic functions for which the pun-target pairs were tagged initially. These tags were then double-checked against the CLAWS tags (see Appendix A) and adjusted. It should be mentioned again, that the present study does not strive for an exhaustive syntactic analysis that is reflected in these tags, but a useable set that can easily be translated for an application to generate puns and that yields a result with respect to the hypothesis sketched above and refined below.

The following statistics emerge from the material on the basis of my hypothesis and the tagging classes used: Of the 162 pun-target pairs that do not show the same word class but are grammatical for both the pun and the target, by far the largest group with 65 instances shows pun and target as premodifiers of a noun head. The pun is either an adjective that is paired with a target that is a premodifying noun as in (79), or, more frequently (52 instances), vice versa, as in (80):

---

[48] Further examples include: active vs. passive voice, indirect object vs. *to* prepositional phrase, and the position of the phrasal verb particle (cf. Greenbaum 1980: 308). Such syntactic freedom has been found desirable for the same reason, namely that it affects meaning as little as necessary, in another NLP application, namely natural language watermarking (Atallah *et al.* 2001).

(79) One of the guards at the Tower of London is so in love with his job that his colleagues call him "the brief eater." (35)
brief (adjective) -> *beef* (premodifying noun)

(80) Birds make tweet music. (238)
tweet (premodifying noun) -> *sweet* (adjective)

Other, less significant subgroups pair the same word classes, adjectives and nouns, in different functions. For example, in the position where the existential verb can have a direct object, a predicative adjective can also show up: *She is affected.* vs. *She is a factoid*. Also, nouns functioning as objects can also appear in a sentence-final position, just as adverbs, so that candidates from these two word classes may form pun-target pairs.

Many of the word class differences involve unequal segmentation, leading to their inclusion in this group of word class differences, possibly with additional word class difference only in one of the segments. Again, noun phrases and premodification were prominently involved (29 instances). Single nouns are paired with two coordinated nouns (example (81); 2 instances), with compound nouns and nouns with a determiner or adjective (example (82); 17 instances). Postmodification makes it possible to pair nouns in object position with nouns and subordinate clauses directly adjacent to them (example (83); 10 instances):

(81) The reason they didn't serve pie at the Last Supper was because they had already had piety. (176)
piety -> pie and tea

(82) The old sheep clipper's back ached so much from yesterday's work that he couldn't make it back to the farm today. He had to send his young assistant, the sore shearer's apprentice. (215)
sore shearer -> sorcerer

(83) Is that a crack in my soup bowl or a heresy? (110)
heresy (noun)-> *hair I see* (noun and relative clause)

In sum, the prominence of noun phrases in the corpus, in particular that of nouns and adjectives in premodifying function, warrants a bias of a generator towards inclusion of these in its selection towards an output. Verb phrases with their more intricate syntax play a less significant role in heterophonic puns and should be considered too costly an issue in terms of their full formalization.

6.8 Recoverability through Ungrammaticality

Related to these parameters, but not strictly dependent on them is the status of syntactic grammaticality of the pun and target as embedded into their context. While we can expect grammaticality for both when word class and syntactic function are identical, we should assume where the syntactic function differs, that only under conditions of structural ambiguity the grammaticality of both pun and target can be achieved. These are cases of punning interacting with the factors creating structural ambiguity described by Oaks (1990, 1995), for example the interaction of ellipsis and ambiguous coordination in the following example.

(84) ‘Tis better to have love and lust
Than never to have love at all.
love and lust (coordination of two direct objects) -> *loved and lost* (coordination of two main verbs)

The statistics for grammaticality and formulaicity of the sample corpus of 1182 texts are given in the following table:

Table (10): Grammaticality (gram.) Formulaicity (form.) of Puns and Targets

grammaticality	#	*form. #*	*% gram.*	*% form. of gram.*
1. both pun and target	**861**	**668**	72.84	77.58
2. only target	**223**	**102**	18.87	45.74
3. not syntactically embedded	**127**	**27**	10.74	21.26
4. only the pun	**1**	**1**	0.085	100.00

The main point that these figures illustrate has been introduced above: successful syntactic embedding of both the pun and target is desirable, but not obligatory.

It is also remarkable that the number of targets whose recoverability is enhanced through the use of formulaic speech is lower for the puns in which only those targets are grammatical (second row), namely only for 46% of the cases in contrast to 78% for the standard cases. I assume that the ungrammaticality of the pun triggers the target differently here than in puns in which both pun and target are grammatical. If this assumption is correct, it explains how the puns in which only the targets are grammatical attain recoverability despite the phonological dissimilarity without having to fall back to formulaicity, as in the following word play:

(85) "Grass" reefers to marijuana. (194)
reefers (noun) -> *refers* (verb)

The pun *reefers* as a plural noun is a synonym for marijuana cigarette. It is ungrammatical in the text, helping to trigger the grammatical target verb *refer* despite their significant sound distance created by stress shift. The whole text thus has but one script, SMOKING MARIJUANA, presumably making it sufficiently enjoyable for those who do, but not sufficient to make it a joke (cf. section 2.3). The contrast between the two readings of the text is just that between a grammatical and an ungrammatical sentence,

one of which requires pun-like problem-solving to be recovered, the other being about drugs. If ungrammatical puns trigger the grammatical targets, this is related to Sobkowiak's observation that puns are also "much less frequent in English than their sources [targets]" (1991: 116). It seems easier to trigger a high-frequency, "common" target with a low-frequency, "rare" pun than vice versa.

The majority of the puns that fall into the third class, in which the pun/target is not embedded into a syntactic context, namely 75 (59.05%), are of the *definition* type, which will be discussed in section 6.10 below.

In conclusion, syntactic functions become relevant for punning in those cases, where the pun has a different word class than its target, while both are syntactically grammatical. For the purpose of a pun generator component as envisioned here—namely one stopping short of producing the semantically well-formed joke text, but providing usable pun-target pairs—syntactic functions are a secondary parameter. While it was necessary to confirm this status in the present study, all that a generator component will take into account is the lexical word class of the two members of the pair. This word class constraint will, on the other hand, reflect the insights gained on the basis of the syntactic functions they can fulfill. In other words, if words from several different lexical word classes can fulfill the same syntactic function and therefore appear in the same position, they form a paradigmatic class whose members are, for the purposes of punning, interchangeable and thus possible target-pun pairs.

6.9 Syntactic Ambiguity

The following subsection discusses a different case of exceptional grammaticality, not so much one that is preserved through shared syntactic function, but rather created through a more intricate ambiguity of the syntactic and semantic structure of the text, not just the pun-target pair. The mechanism that creates this ambiguity is the embedding into a sentence whose syntax is itself ambiguous and allows for two syntactic functions and/or word classes for the pun-target pair. A small number of puns in my sample corpus utilized this principle, and it sees worthwhile to investigate these jokes as they struck me as particularly witty or, at least, intricate and thus, maybe, sophisticated. These puns require additional syntactic resources for their analysis than those outlined so far. Fortunately, Oaks has devoted not only his dissertation (1990), but also a number of subsequent publications (e.g., 1995) to precisely these types of syntactic phenomena that enable ambiguity.

So far, syntactic factors have been discussed in their function as disambiguators, in particular when they helped trigger the recovery of the target in a context in which the pun was ungrammatical. Focusing on the opposite function, this section will summarize Oak's discussion briefly insofar as it is required to account for the material found in the sample corpus. The distinct quality of ambiguity discussed in this subsection becomes clear when it is contrasted to the similar, but simpler cases already discussed:

(86) The secret of being a happy woman over forty is not to worry about the menopause but the men who don't. (148)
menopause -> man who paws, PDetPo -> PDetPoWhV[49]

[49] For an explanation of the abbreviations of word classes, see Appendix B.

The word class in this example is the same for the pun and the head of the phrase onto which it is projected as the target, namely a noun phrase functioning as the object, with an additional relative clause in the case of the target.

(87) To do a tapestry, you needle the wool you can get. (158)
needle -> need all, SVDetO -> SVDetDetO

Again, we have two different verb phrases, depending on how the punning domain is segmented, in case of the target providing an additional premodifying quantifier *all*. But these aren't examples of ambiguous syntactic patterns. It is merely the function of a coextensive segment of the joke that has two different possible segmentations, but its syntactic function, its position within the hierarchical phrase structure, is still the same in the pun and target.

(88) The Spartan mothers used to enjoin their sons: "Return with your shield—or pawn it." (172)
pawn -> upon, VPAdjPo&PPo ->VPAdjPo&VO, Po&Po ->V&V

This example of true structural ambiguity employs the potential of coordination and ellipsis, the combination of which yields a rich harvest of syntactic patterns that can be understood in two ways (cf. Oaks 1990: 117ff). Of this latter type, I encountered 23 in the sample corpus, mostly employing coordination of elliptical phrases. The other cases involved different types of pre- and postmodification of NPs and VPs between the target and pun and are subsumed under the constraints on word class interchangeability discussed in the previous section.

The following are further examples of punning with additional structural ambiguity, largely facilitated by coordination and ellipsis in connection.

(89) The poet became unable to scan anapest. (13)
anapest -> and a pest, PdjToVO->PdjTo&DetO coordination/ellipsis

(90) Go to Lebanon and cedar Lebanon. (44)
cedar -> see the, VPPo&RPo->VPPo&VDetO coordination/ellipsis

(91) You can't live by bread and loaf. (140)
and loaf -> alone, SAuxVPPo&Po/V[50] -> SAuxVPPoAdv (coordination)

(92) She spent her time
A lot alone
And never sad
Upon the throne. (201)
sad -> sat, SVOAdv&AdvPDetPo ->SVOAdv&VPDetPo coordination/ellipsis.

6.10 Definition Puns: Less Semantics, Less Joke

Let me finally turn to one statistically significant class of puns, whose syntactic grammaticality and, consequently, syntactic function and lexical word class are irrelevant. A total of 92 of the 1182 puns (7.78%) in the sample corpus are of this "definition" type, most of them neither embedding the pun nor the target syntactically (see above). Rather, the pun is presented unembedded into context in the fashion of a lexicon entry, and then just defined such that the target can be arrived at from the second script triggered by the definition. Often these puns are so semantically ill-formed that to call them jokes is actually straining the category to the point of breaking (see, again, section 2.3). In the following example, the second script triggered by the definition (*itch*), namely SKIN IRRITATION, is only marginally overlapping with weak first script of SMALLNESS (triggered by a grammatical lexeme *each*) as it should be provided by the meaning of the unembedded pun:

[50] I'm grateful to Shaun Hughes who pointed out to me that within the pun, *loaf* is also ambiguous as a noun denoting "bread loaf" thus reinforcing the FOOD script, and—more plausible in the context—as a verb denoting "pass time at leisure."

(93) each
minor irritation (73)
each -> itch

The meaning of *each* is only marginally supporting the SMALLNESS of the irritation, in that it denotes single iteration or instance of whatever it is the quantifier of. Thus it isn't as bad a joke as many other definition puns, and should not be considered purely wordplay, in contrast to the following one. Although this example has word-class identity for pun and target, I challenge the reader to unpack a logical mechanism (LM) creating overlap beyond the sound similarity or identify the scripts behind that could connect *maggot* to *magnet* in this joke-like text.

(94) magnet
To some, it is what you find in a bad apple. (145)
magnet -> maggot

These definition puns are also an extreme in terms of the interaction of semantic, phonological, and syntactic interaction towards recoverability of the target. There is zero syntactic support for the target, hence the semantic and phonological forces must be either maximal or the text will fail as a joke, despite its success in making a target recoverable from a pun (see section 2.3). And this illustrates another assumption we can derive from the theoretical model outlined in that section: Because the LM of punning is the analogical extension of cratylism, if there are too many supporting factors compensating for low phonological similarity, this LM is negatively affected, despite the recoverability of the target. Since we assume the LM to be an essential element in the joke, its functioning—responsible for its funniness potential—should be affected. And I argue that it indeed is.

7. SUMMARY AND OUTLOOK

In the present study, a model for the complex phenomenon of punning, and heterophonic punning, in particular has been described, at all levels of linguistic and humor theoretical relevance. The level of formalization that was attainable for the different domains that this model pertains to depends strongly on the state of the art of the particular theory underlying it, as well as to what degree it was possible to advance that state here. Thus, it was possible to provide more detail to the humor theoretical model of punning, most crucially from a semantic perspective, and describe its intricate interactions with other linguistically-relevant factors, in particular the phonological concept of sound similarity as it operates in a specific way in imperfect punning and interacts with other linguistic factors. It was also possible to evaluate the limits of Optimality Theory (OT) with a view to its application and extend models that overcome these limits while retaining the essential advantages of an underlying theory inspired by OT. Further significant results for phonology include a feature hierarchy based on the assumed influence of different features and feature classes on perceived similarity. In this view, the present study provides relevant insights beyond the data in its purview, namely for those theories that were used and adapted to provide models for the description and explanation of those data.

From the opposite perspective of application and implementation, the complexity and multifaceted nature of the phenomenon of punning itself provides a criterion with which the level of abstraction—or, inversely, detail—of the model can be decided. In particular, the implementation of the current model for computational generation demands the full description of the relevant aspects, namely semantics and phonology to produce the promised output, namely ranked selection of pun candidates for an input target. For this reason the focus of the present study was directed towards the development of present theories in these areas towards the description of imperfect punning.

The phonological analysis of possible puns for a given target can be automatized and refined as part of a generator, now that the edit distance computation has been made available by the present research. This method to evaluate and select phonologically possible and better imperfect puns can thus be integrated into larger natural language processing project as a module to generate imperfect puns. Future extensions and applications of the results presented here in this direction will also lead to further refinement of the violation cost calculation presented in section 5.7. Such a module can function as a complement to a existing automatized pun generator (cf. Binsted and Ritchie 1997, Stock and Strapparava 2002) that are currently only capable to produce homophonic, perfect puns.

BIBLIOGRAPHY

As customary, papers available in the Rutgers Optimality Archive (ROA) at <http://roa.rutgers.edu> are cited with their ROA number.

Ad C. Herennium. De Ratione Dicendi (Rhetorica ad Herennium). Trans. by Harry Caplan. Cicero in Twenty Eight Volumes. Vol. I. Cambridge, MA: Harvard University Press, London: Heinemann, 1981.

Albro, Daniel M. 2003. A Large-Scale, Computerized Phonological Analysis of Malagasy. Paper presented at the 2003 LSA Meeting, Atlanta, GA. available at <http://www.humnet.ucla.edu/people/albro/albro.htm>.

Allan, Keith. 2001. "Sound Symbolism." In: *Natural Language Semantics.* London: Blackwell: 132-141.

Apte, Mahadev. 1987. "Introduction." *International Journal of the Sociology of Language* 65: 5-8.

Archangeli, Diana. 1996. "Optimality Theory: An Introduction to Linguistics in the 1990s." In: Archangeli, Diana, and D. Terence Langendoen. Eds. 1996: 1-32.

Archangeli, Diana, and D. Terence Langendoen. Eds. 1996. *Optimality Theory: An Overview.* London: Blackwell.

Atallah, Mikhail J., Victor Raskin, Michael Crogan, Christian F. Hempelmann, Florian Kerschbaum, Dina Mohamed, and Sanket Naik. 2001. "Natural Language Watermarking: Design, Analysis, and a Proof-of-concept Implementation." In: Moskowitz, Ira S. Ed. *Information Hiding. Proceedings of the Fourth International Workshop IH 2001.* Berlin, New York: Springer: 185-199.

Attardo, Salvatore. 1994. *Linguistic Theories of Humor.* Berlin: Mouton de Gruyter.

Attardo, Salvatore. 1997. "The Semantic Foundations of Cognitive Theories of Humor." *HUMOR* 10-4: 395-420.

Attardo, Salvatore. 2001. *Humorous Texts: A Semantic and Pragmatic Analysis.* Berlin: Mouton de Gruyter.

Attardo, Salvatore and Victor Raskin. 1991. "Script Theory Revis(it)ed: Joke Similarity and Joke Representation Model." *HUMOR* 4:3-4: 293-347.

Attardo, Salvatore, Christian F. Hempelmann, and Sara Di Maio. 2002. "Script oppositions and logical mechanisms: Modeling incongruities and their resolutions." *HUMOR* 15-1: 3-46.

Auden, W.H. 1976. "Natural Linguistics." *Collected Poems.* New York: Random House.

Barnes, Gary James, Harriet G. Kopp, and Alan C. Nichols.1981. "Effects of Phonetic Distance on Recognition of Synthetic Sentences." *Journal of Communication Disorders* 14: 341-47.

Baum, Paull F. 1956. "Chaucer's Puns." *PMLA* 71: 225-46.

Benua, Laura. 1997. *Transderivational Identity: Phonological Relations Between Words.* PhD Thesis. Amherst: UMass. ROA 259.

Binsted, Kim. 1995. *Using humour to make natural language interfaces more friendly.* Unpublished Ms. University of Edinburgh.

Binsted, Kim and Graeme Ritchie. 1994. "An implemented model of punning riddles." In: *Proceedings of the Twelfth national Conference on Artificial Intelligence (AAAI-94).* Seattle, USA.

Binsted, Kim and Graeme Ritchie. 1997. "Computational rules for generating punning riddles." *HUMOR* 10-1: 25-76.

Bloomfield, Leonard. 1895. "On Assimilation and Adaptation in Congeneric Classes of Words." *American Journal of Philology* 16: 409-34.

Bolinger, Dwight. 1965. "The Atomization of Meaning." *Language* 41: 555-73.

Bolinger, Dwight. 1971. *The Phrasal Verb in English.* Cambridge, MA: Harvard University Press.

Braune, Wilhelm. 1966. *Gotische Grammatik*. Ed. Ernst A. Ebbinghaus. Tübingen: Niemeyer.

Brinton, Laurel J. 1988. *The Development of English Aspectual Systems.* Cambridge: Cambridge University Press.

Cassirer, Ernst. 1946. *Language and Myth.* Trans. Susanne K. Langer. New York, London: Harper.

Chaucer, Geoffrey. 1987. *The Riverside Chaucer*. Ed. Larry D. Benson. Boston: Houghton Mifflin.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper and Row.

Cerf, Bennett. 1968. *Treasury of Atrocious Puns*. New York: Harper and Row.

Clements, G. N. and Elizabeth V. Hume. 1995. "The Internal Organization of Speech Sounds." In: Goldsmith, John A. Ed. *Handbook of Phonological Theory*. London: Blackwell: 245-306.

Coleman, John and Janet Pierrehumbert. 1997. *Stochastic Phonological Grammars and Acceptability*. Proceedings of SIGPHON 3.

Coleridge, Samuel Taylor. 1907. *Biographia Literaria.* London: Oxford University Press.

Comrie, Bernard. 1979: "Morphophonemic Exceptions and Phonetic Distance." *Linguistics* 18: 51-60.

Crosbie, John S. 1977. *Crosbie's Dictionary of Puns.* New York: Harmony.

Cutler, Anne and Charles Clifton Jr. 1999. "Comprehending spoken language: A blueprint of the listener." In: Brown, C.M. and P. Hagoort. *The Neurocognition of Language*. London: Oxford University Press: 123-166.

Dale, Robert, Barbara Di Eugenio, and Donia Scott. 1998. "Introduction to the Special Issue on Natural Language Generation." *Computational Linguistics* 24-3: 345-53.

Dalrymple, Mary, R. Kaplan, L. Karttunen, M. Kay, A. Kornai, K. Koskenniemi, S. Shaio, and M. Wescoat. 1987. *DKIMMO/TWOL: A Development Environment for Morphological Analysis.* Stanford, CA: Xerox Palo Alto Research Center and Center for the Study of Language and Information.

Eco, Umberto. 1983. "Horns, Hooves, Insteps: Some Hypotheses on Three Types of Abduction." In: Eco, Umberto and Sebeok, Thomas A. Eds. *The Sign of the Three.* Bloomington: Indiana University Press: 198-220.

Ellison. Mark T. 1995. "Phonological Derivation in Optimality Theory." *Proceedings of the 15th International Conference on Computational Linguistics*: 1007-13. ROA 75.

Erhard Voeltz, F. K. and Christa Kilian-Hatz. Eds. 2001. *Ideophones*. Amsterdam: Benjamins.

Fleischhacker, Heidi. 2002. *Onset Transfer in Reduplication*. Paper presented at the LSA meeting, San Francisco, January 2002. Available at: http://www.linguistics.ucla.edu/people/grads/fleischhacker/2002lsapaper.pdf.

Frazer, James George. 1935. *The Golden Bough: A Study in Magic and Religion.* Part II. *Taboo and the Perils of the Soul.* New York: Macmillan.

Freud, Sigmund. 1950. *Totem and Taboo: Some Points of Agreement between the Mental Lives of Savages and Neurotics.* Trans. James Strachey. New York, London: Norton.

Freud, Sigmund. 1960. *Jokes and their Relation to the Unconscious.* Trans. James Strachey. New York, London: Norton.

Frisch, Stefan. 1996. *Similarity and Frequency in Phonology*. PhD Thesis. Evanston: Northwestern University. ROA 198.

Fromkin, Victoria. Ed. 1973. *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Fromkin, Victoria. Ed. 1980. *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic.

Gallo, Ernest. 1971. *The* Poetria Nova *and its Sources in Early Rhetorical Doctrine.* The Hague, Paris: Mouton.

Garside, Roger. 1997. *Using CLAWS to annotate the British National Corpus.* Department of Computing, University of Lancaster. Available at: http://www.hcu.ox.ac.uk/BNC/what/garside_allc.html. Thu Sep 11 07:05:22 1997.

Geoffrey of Vinsauf. 1967. *Poetria nova of Geoffrey of Vinsauf.* Trans. Margaret F. Nims. Toronto: Pontifical Institute of Mediaeval Studies.

Golston, Chris. 1995. *Direct Optimality Theory: Representation as Constraint Violation.* ROA-71.

Greenbaum, Sidney. 1980. "Syntactic Frequency and Acceptability." In: Thomas Perry. Ed. *Evidence and Argumentation in Linguistics*. Berlin, New York: de Gruyter: 301-14.

Hale, Mark, Madelyn Kissock, and Charles Reiss. 1998. "Output-Output Correspondence in Optimality Theory." In: *The Proceedings of the 16th West Coast Conference on Formal Linguistics.* Curtis, Emily, Lyle, James, & Webster, Gabriel. Eds. Stanford, CA: Center for the Study of Language & Information: 223-36.

Hall, Patrick A. V. and Geoff R. Dowling. 1980. "Approximate String Matching." *ACM Computing Surveys* 12-4: 381-402.

Hammond, Michael. 1997. *Parsing Syllables: Modeling OT Computationally.* ROA 222.

Hausmann, Franz Josef. 1974. *Studien zur Linguistik des Wortspiels.* Tübingen: Niemeyer.

Hempelmann, Christian F. 2003. " '99 nuns giggle, 1 nun gasps:' The not-all-that-Christian Natural Class of Christian Jokes." *HUMOR*. 16-1: 1-31.

Hinton, Leanne, Johanna Nichols and John Ohala. 1994. "Introduction: Sound-Symbolic Processes." In: Hinton *et al*. Eds.: 1-12.

Hinton, Leanne, Johanna Nichols, and John Ohala. Eds. 1994. *Sound Symbolism.* Cambridge, Cambridge University Press.

Hockett, Charles. 1958. *A Course in Modern Linguistics.* New York: Macmillan.

Hooper, Joan B. 1976. "Constraints on schwa-deletion in American English." In: *Recent Developments in Historical Phonology*. Jacek Fisiak. Ed. The Hague: Mouton: 183-207.

Hulstijn, Joris and Anton Nijholt. Eds. 1996. *Automatic Interpretation and Generation of Humor. Proceedings Twente Workshop on Language Technology 12*. Enschede: Universiteit Twente.

Humboldt, Wilhelm. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts.* Berlin: Königliche Akademie.

Humboldt, Wilhelm. 1971. *Linguistic Variability and Intellectual Development.* Trans. George C. Buck and Fritjof A. Raven. Coral Gable, FL: University of Miami Press.

Hume, Elizabeth. 2001. "Metathesis: Formal and Functional Considerations." In: Hume, Elizabeth, Norval Smith, and Jeroen van de Weijer. Eds. *Surface Syllable Structure and Segment Sequencing.* HIL Occasional Papers. Leiden: HIL: 1-25.

Iverson, Gregory K. 1995. "Rule Ordering." In: Goldsmith, John A. Ed. *Handbook of Phonological Theory*. London: Blackwell: 609-14.

Joanette, Yves, Erich Keller, and André Roch Lecours. 1980. "Sequences of Phonemic Approximation in Aphasia." *Brain and Language* 11: 30-44.

Jakobson, Roman and Morris Halle. 1956. *Fundamentals of Language*. s'Gravenhage: Mouton.

Jakobson, Roman and Linda Waugh. 1979. *The Sound Shape of Language*. Bloomington: Indiana University Press.

Kapnick, Sharon. 2003. "America: Land of the Red, White and Rosé." *Time* 13. Mar.: 83.

Karttunen, Lauri. 1983. "KIMMO: a general morphological processor." *Texas Linguistic Forum* 22:163-86.

Karttunen, Lauri. 1998. "The Proper Treatment of Optimality in Computational Phonology." In: *Proceedings of FSMNLP'98. International Workshop on Finite-State Methods in Natural Language Processing.* Ankara: Bilkent University: 1-12. ROA 258.

Kenstowicz, Michael. 1994a. *Phonology in Generative Grammar*. Oxford: Blackwell.

Kenstowicz, Michael. 1994b. *Base-Identity and Uniform Exponence: Alternatives to Cyclicity.* ROA 103.

Knott, Eleanor. 1974. *An Introduction to Irish Syllabic Poetry of the Period 1200-1600.* Dublin: Dublin Institute for Advanced Studies.

Köhler, Wolfgang. 1929. *Gestalt Psychology.* New York: Liveright.

Ladefoged, Peter. 2001. *Vowels and Consonants. An Introduction to the Sounds of Languages.* Oxford: Blackwell.

Lagerquist, Linnes M. 1980. "Linguistic Evidence from Paronomasia." *CLS* 16. 185-91.

Loehr, Dan. 1996. "An Attempt at Natural Humor from a Natural Language Robot." In: Hulstijn and Nijholt. Eds.: 161-71.

Lutz, Richard and Stephan Greene. 2001. *Measuring Phonological Similarity: The Case of Personal Names.* Herndon, VA: Language Analysis Systems, Inc.

Makkai, Adam. 1972. *Idiom Structure in English.* The Hague, Paris: Mouton.

"Man has Derogatory Nicknames for Every Neighboring Town." *The Onion.* 26 Feb. 2003.

Marchand, Hans. 1957. "Phonetic Symbolism in English Word-Formation. Part I." *Indogermanische Forschungen* 63:146-68.

Marchand, Hans. 1958. "Phonetic Symbolism in English Word-Formation. Part II." *Indogermanische Forschungen* 64: 256-77.

Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word-Formation.* München: Beck.

Mayerthaler, W. 1982. "Markiertheit in der Phonologie." In: Vennemann, T. Ed. *Silben, Segmente, Akzente*. Tübingen: Narr. 205-46.

McCarthy, John J. 1996. *Faithfulness in Prosodic Morphology & Phonology: Rotuman Revisited.* ROA 110.

McCarthy, John J. and Alan S. Prince. 1995. *Faithfulness and Reduplicative Identity.* ROA 60.

McDonough, Craig J. 2001. *Mnemonic String Generator: Software to aid memory of random passwords.* CERIAS Technical Report. West Lafayette, IN: Purdue University.

McGregor, William B. 1996. "Sound Symbolism in Gooniyandi, a Language of Western Australia." *Word* 47-3: 339-64.

Mohr, B. and W. S.-Y. Wang. 1968. "Perceptual Distance and the Specification of Phonological Features." *Phonetics* 18: 31-45.

Monnot, Michel. 1981. *Selling America - Puns, Language, and Advertising.* Lanham: University Press of America.

Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach.* Oxford: Clarendon.

Morkes, John, Hadyn K. Kernal, and Clifford Nass. 1998. "Humor in Task-Oriented Computer-Mediated Communication and Human-Computer Interaction." *Proceedings of CHI 98*. New York: ACM: 215-16.

Morris, Charles. W. 1946. *Sign, Language, and Behavior.* New York: Braziller.

Mulder, Matthijs P. and Anton Nijholt, Anton. 2002. *Humour Research: State of the Art*. Technical Report. Enschede: Universiteit Twente.

Nara. 1995. *Using the Census Soundex.* General Information Leaflet 55.

Niepokuj, Mary. 1997. *The Development of Verbal Reduplication in Indo-European.* Journal European Studies Monograph 24. Washington, DC: Institute for the Study of Man.

Nijholt, Anton. 2002. *Embodied Agents: A New Impetus to Humor Research.* In: Stock, Oliviero, Carlo Strapparava, and Anton Nijholt. Eds.: 101-11.

Nirenburg, Sergei and Victor Raskin. 2003. *Ontological Semantics.* Cambridge, MA: MIT Press.

Nuckolls, Janis B. 1996. *Sounds like Life: Sound-Symbolic Grammar, Performance, and Cognition in Pastaza Quechua.* New York and Oxford: Oxford University Press.

Oaks, Dallin D. 1990. *Enablers of Grammatical Ambiguity*. Dissertation: Purdue University.

Oaks, Dallin D. 1995. "Structural Ambiguities and Written Advertisements: An Inventory of Tools for More Resourceful Advertisements in English." *Journal of Technical Writing and Communication* 25-4: 371-92.

Pepicello, William J. and Thomas A. Green. 1984. *The Language of Riddles: New Perspectives.* Columbus, OH: Ohio State University Press.

Pepicello, William J. and Rover W. Weisberg. 1983. "Linguistics and Humor." In: McGhee, Paul E. And Jeffrey H. Goldstein. Eds. *Handbook of Humor Research.* 2 Vols. New York, Berlin, Heidelberg, Tokyo: Springer.

Philips, Lawrence. 1990. "Hanging on the Metaphone." *Computer Language* 7-12: 39-43.

Plato. 1961. *Collected Dialogues.* Ed. Edith Hamilton and Huntington Cairns. Princeton, NJ: PUP.

Preisendanz, Wolfgang. 1970. *Über den Witz.* Konstanz: Konstanzer Universitätsreden 13.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. ROA 537.

Pulleyblank, Douglas. 1996. "Optimality Theory and Features." In: Archangeli, Diana, and D. Terence Langendoen. Eds.: 59-101.

Raskin, Victor. 1985. *Semantic Mechanisms of Humor.* Dordrecht: Reidel.

Raskin, Victor. 1996. "Computer Implementation of the General Theory of Verbal Humor." in Hulstijn, Joris and Anton Nijholt. Eds.: 9-19.

Raskin, Victor. 2002. Quo Vadis *Computational Humor?* In: Stock, Oliviero, Carlo Strapparava, and Anton Nijholt. Eds.: 31-46.

Raskin, Victor and Salvatore Attardo. 1994. "Non-literalness and non-bona-fide in language: An approach to formal and computational treatments of humor." *Pragmatics and Cognition* 2-1: 31-69.

Redfern, Walter. 1984. *Puns*. London: Blackwell.

Ritchie, Graeme. 2001. "Current Directions in Computational Humour." *Artificial Intelligence Review* 16: 119-35.

Ritchie, Graeme D., Graham J. Russell, Alan W. Black, and Stephen G. Pulman. 1992. *Computational Morphology: Practical Mechanisms for the English Lexicon.* Cambridge, MA: MIT Press.

Robinson, F.N. Ed. 1957. *The Works of Geoffrey Chaucer.* 2nd ed. Boston, MA: Houghton Mifflin.

Ruch, Willibald, Salvatore Attardo, and Victor Raskin. 1993. "Toward an Empirical Verification of the General Theory of Verbal Humor." *HUMOR* 6-2: 123-36.

de Saussure, Ferdinand. 1983. *Course in General Linguistics*. Trans. Roy Harris. Chicago and La Salle, IL: Open Court.

Sebeok, Thomas A. 1976. *Contributions to the Doctrine of Signs.* Bloomington, IN: Indiana University Press.

Sobkowiak, Wlodzimierz. 1991. *Metaphonology of English Paronomasic Puns*. Frankfurt: Lang.

Stanford, W.B. 1972. *Ambiguity in Greek Literature*. New York: Johnson.

Steriade, Donca. 2001. "The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization." UCLA. Ms. Available at http://www.humnet.ucla.edu/linguistics/people/steriade/steriade.htm.

Steriade, Donca and Jie Zhang. 2001 "Context-Dependent Similarity: Phonology & Phonetics of Romanian Semi Rimes." Presentation at CLS 37.

Stock, Oliviero and Carlo Strapparava. 2002. *HAHAcronym: Humorous Agents for Humorous Acronyms.* In: Stock, Oliviero, Carlo Strapparava, and Anton Nijholt. Eds.: 125-35.

Stock, Oliviero, Carlo Strapparava, and Anton Nijholt. Eds. 2002. *TWLT20: The April Fools' Day Workshop on Computational Humor.* Enschede: Universiteit Twente.

Taylor, W.L. 1953. " 'CLOZE' Procedure: A New Tool for Measuring Readability." *Journalism Quarterly* 30: 415-33.

Tesar, Bruce. 1995. *Computational Optimality Theory.* Ph.D. Thesis. Boulder, CO: University of Colorado.

Trudgill, Peter. 1990. *The Dialects of England.* Oxford: Blackwell.

Vitz, Paul C. and Brenda Spiegel Winkler. 1973. "Predicting the Judged 'Similarity of Sound' of English Words." *Journal of Verbal Learning and Verbal Behavior* 12: 373-88.

Zwicky, Arnold M. 1976. "Well, this rock and roll has got to stop. Junior's head is as hard as a rock." *Papers from the CLS Regional Meetings* 12: 676-97.

Zwicky, Arnold M. and Elizabeth D. Zwicky. 1986. "Imperfect Puns, Markedness, and Phonological Similarity: With Fronds like These, Who Needs Anemones?" *Folia Linguistica* 20-2: 493-503.

APPENDICES

Appendix A: CLAWS Tag Set

The following is the CLAWS tag set from Garside (1997) that was used for the word-class identification.

AJ0 Adjective (general or positive) (e.g. *good, old, beautiful*)

AJC Comparative adjective (e.g. *better, older*)

AJS Superlative adjective (e.g. *best, oldest*)

AT0 Article (e.g. *the, a, an*, *no*) [N.B. *no* is included among articles, which are defined here as determiner words which typically begin a noun phrase, but which cannot occur as the head of a noun phrase.]

AV0 General adverb: an adverb not subclassified as AVP or AVQ (see below) (e.g. *often, well, longer* (adv.), *furthest*. [Note that adverbs, unlike adjectives, are not tagged as positive, comparative, or superlative. This is because of the relative rarity of comparative and superlative adverbs.]

AVP Adverb particle (e.g. *up, off, out*) [N.B. AVP is used for such "prepositional adverbs", whether or not they are used idiomatically in a phrasal verb: e.g. in 'Come *out* here' and 'I can't hold *out* any longer', the same AVP tag is used for *out*.

AVQ *Wh*-adverb (e.g. *when, where, how, why, wherever*) [The same tag is used, whether the word occurs in interrogative or relative use.]

CJC Coordinating conjunction (e.g. *and, or, but*)

CJS Subordinating conjunction (e.g. *although, when*)

CJT The subordinating conjunction *that* [N.B. *that* is tagged CJT when it introduces not only a nominal clause, but also a relative clause, as in 'the day *that* follows Christmas'. Some theories treat *that* here as a relative pronoun, whereas others treat it as a conjunction. We have adopted the latter analysis.]

CRD Cardinal number (e.g. *one, 3, fifty-five, 3609*)

DPS Possessive determiner (e.g. *your, their, his*)

DT0 General determiner: i.e. a determiner which is not a DTQ. [Here a determiner is defined as a word which typically occurs either as the first word in a noun phrase, or as the head of a noun phrase. E.g. *This* is tagged DT0 both in '*This* is my house' and in '*This* house is mine'.]

DTQ *Wh*-determiner (e.g. *which, what, whose, whichever*) [The category of determiner here is defined as for DT0 above. These words are tagged as *wh*-determiners whether they occur in interrogative use or in relative use.]

ITJ Interjection or other isolate (e.g. *oh, yes, mhm, wow*)

NN0 Common noun, neutral for number (e.g. *aircraft, data, committee*) [N.B. Singular collective nouns such as *committee* and *team* are tagged NN0, on the grounds that they are capable of taking singular or plural agreement with the following verb: e.g. 'The *committee* disagrees/disagree'.]

NN1 Singular common noun (e.g. *pencil, goose, time, revelation*)

NN2 Plural common noun (e.g. *pencils, geese, times, revelations*)

NP0 Proper noun (e.g. *London, Michael, Mars*, *IBM*) [N.B. the distinction between singular and plural proper nouns is not indicated in the tagset, plural proper nouns being a comparative rarity.]

ORD Ordinal numeral (e.g. *first, sixth, 77th, last*) . [N.B. The ORD tag is used whether these words are used in a nominal or in an adverbial role. *Next* and *last,* as "general ordinals", are also assigned to this category.]

PNI Indefinite pronoun (e.g. *none, everything, one* [as pronoun], *nobody*) [N.B. This tag applies to words which always function as [heads of] noun phrases. Words like *some* and *these*, which can also occur before a noun head in an article-like function, are tagged as determiners (see DT0 and AT0 above).]

PNP Personal pronoun (e.g. *I, you, them, ours*) [Note that possessive pronouns like *ours* and *theirs* are tagged as personal pronouns.]

PNQ *Wh*-pronoun (e.g. *who, whoever, whom*) [N.B. These words are tagged as *wh*-pronouns whether they occur in interrogative or in relative use.]

PNX Reflexive pronoun (e.g. *myself, yourself, itself, ourselves*)

POS The possessive or genitive marker *'s* or *'* (e.g. for 'Peter*'s* or somebody else*'s*', the sequence of tags is: NP0 POS CJC PNI AV0 POS)

PRF The preposition *of*. Because of its frequency and its almost exclusively postnominal function, *of* is assigned a special tag of its own.

PRP Preposition (except for *of*) (e.g. *about, at, in, on, on behalf of, with*)

TO0 Infinitive marker *to*

UNC Unclassified items which are not appropriately classified as items of the English lexicon. [Items tagged UNC include foreign (non-English) words, special typographical symbols, formulae, and (in spoken language) hesitation fillers such as *er* and *erm*.]

VBB The present tense forms of the verb BE, except for *is, 's*: i.e. *am, are, 'm, 're* and *be* [subjunctive or imperative]

VBD The past tense forms of the verb BE: *was* and *were*

VBG The *-ing* form of the verb BE: *being*

VBI The infinitive form of the verb BE: *be*

VBN	The past participle form of the verb BE: *been*
VBZ	The -*s* form of the verb BE: *is, 's*
VDB	The finite base form of the verb BE: *do*
VDD	The past tense form of the verb DO: *did*
VDG	The -*ing* form of the verb DO: *doing*
VDI	The infinitive form of the verb DO: *do*
VDN	The past participle form of the verb DO: *done*
VDZ	The -*s* form of the verb DO: *does, 's*
VHB	The finite base form of the verb HAVE: *have, 've*
VHD	The past tense form of the verb HAVE: *had, 'd*
VHG	The -*ing* form of the verb HAVE: *having*
VHI	The infinitive form of the verb HAVE: *have*
VHN	The past participle form of the verb HAVE: *had*
VHZ	The -*s* form of the verb HAVE: *has, 's*
VM0	Modal auxiliary verb (e.g. *will, would, can, could*, *'ll*, *'d*)
VVB	The finite base form of lexical verbs (e.g. *forget, send, live, return*) [Including the imperative and present subjunctive]
VVD	The past tense form of lexical verbs (e.g. *forgot, sent, lived, returned*)
VVG	The -*ing* form of lexical verbs (e.g. *forgetting, sending, living, returning*)
VVI	The infinitive form of lexical verbs (e.g. *forget, send, live, return*)
VVN	The past participle form of lexical verbs (e.g. *forgotten, sent, lived, returned*)
VVZ	The -*s* form of lexical verbs (e.g. *forgets, sends, lives, returns*)
XX0	The negative particle *not* or *n't*
ZZ0	Alphabetical symbols (e.g. *A, a, B, b, c, d*)

Appendix B: Syntactic Function Tag Set

nouns

N	not syntactically embedded noun, or ungrammatically used (usually in the pun)
S	subject
O	object
Po	object of a preposition
R	premodifying noun in a compound noun, either spelled together, hyphenated or in two words
C	noun that is a comparative complement
PropX	e.g., PropN, PropR; proper name [if I'm distinguishing this, should I do morphological analysis in terms of deverbal noun, deadjectival noun, etc.? Yes, in the long run, if I have a full pun generator in mind and I can't fit a verb where I need a certain stem semantically, I can adjectivalize it, nominalize it, etc. as needed; but that later]
Rposs	possessive premodifying noun
Wh	interrogative/relative pronoun

adjectives

Adj	attributive adjective
Pdj	predicative adjective

verbs[51]

V	finite verb
Ven	past participle
Ving	present participle
Aux	auxiliary
Inf	infinitive

other

Adv	adverbial
&	conjunction
P	preposition
excl.	exclamation (not syntactically embedded)
Pt	particle of phrasal verbs
to	to (Inf)
Det	determiner
?	indeterminable

[51] Further distinctions are available at the level of the CLAWS-based word class analysis, so that these three V types were only used to highlight central differences.

Appendix C: Consonant Distributions and Frequency

The following table lists the frequencies of consonants in three different positions as extracted from the 127,070 entries of the CMUDict: word-initial (x_), word-final (_x) and word-medial (_x_). These figures were used to assess if the consonants were subject to change in certain environments overproportionately to their regular occurrence.

[b]		
x_	9,087	
_x	416	
_x_	10,403	
sum	*19,906*	
[p]		
x_	7,377	
_x	1,075	
_x_	10,108	
sum	*18,560*	
[d]		
x_	7,452	
_x	8,149	
_x_	14,918	
sum	*30,519*	
[t]		
x_	5,266	
_x	8,890	
_x_	29,613	
sum	*32,050*	
[g]		
x_	4,794	
_x	1.032	
_x_	7,085	
sum	*12,911*	
[k]		
x_	12,412	
_x	5,125	
_x_	22,845	
sum	*40,382*	
[v]		
x_	2,322	
_x	824	
_x_	7,034	
sum	*10,180*	
[f]		
x_	5,294	
_x	969	
_x_	6,673	
sum	*12,936*	
[z]		
x_	918	
_x	18,810	
_x_	6,019	
sum	*25,747*	
[s]		
x_	11,877	
_x	11,035	
_x_	24,571	
sum	*47,483*	
[ʃ]		
x_	2,369	
_x	685	
_x_	5,197	
sum	*8,251*	
[ʒ]		
x_	85	
_x	21	
_x_	431	
sum	*537*	
[θ]		
x_	612	
_x	637	
_x_	1,518	
sum	*2,767*	
[ð]		
x_	65	
_x	40	
_x_	450	
sum	*555*	
[tʃ]		
x_	1,195	
_x	611	
_x_	2,896	
sum	*4,702*	
[dʒ]		
x_	2,039	
_x	634	
_x_	3,259	
sum	*5,932*	
[h]		
x_	6,438	
_x	26	
_x_	2,446	
sum	*8,910*	

[w]

x_	3,659
_x	5
_x_	4,860
sum	*8,524*

[m]

x_	9,029
_x	2,476
_x_	16,433
sum	*27,938*

[l]

x_	5,301
_x	7,153
_x_	34,763
sum	*47,217*

[n]

x_	3,068
_x	14,402
_x_	40,221
sum	*57,691*

[r]

x_	7,107
_x	1,747
_x_	34,763
sum	*43,619*

[ŋ]

x_	0
_x	5,526
_x_	3,842
sum	*9,368*

[j]

x_	1,277
_x	7
_x_	3,523
sum	*4,807*

Appendix D: Non-Corresponding Segments

The following table lists the non-corresponding segments in the main corpus of target-pun pairs listed in Appendix E, including the number of the pair in the database, the pair itself, the immediate segmental environment, and the cost incurred by the non-correspondence violation according to the preliminary cost table. The leading numbers correspond to those in Appendix E, where the full target-pun domain of the examples is listed.

1	z-s	n_#	0.07	26	E-7	C_r	0.29	47	.-t	._#	0.07
2	E-i	n_D	0.28	27	C-S	#_i	0.15	48	l-r	f_4	0.13
3	.-n	0_l	0.1	28	1-3	k_n	0.28	49	.-r	f_a	0.03
3	n-f	5_#	0.66	29	E-a	k_m	0.27	50	C-f	#_.	1
4	I-^	d_n	0.29	30	n-.	0_t	0.1	50	.-r	C_5	0.03
5	&-a	k_m	0.18	30	t-k	n_r	0.14	51	r-.	f_0	0.04
5	l-.	0_#	0.08	31	&-0	k_n	0.24	51	0-^	r_m	0.28
6	s-T	I_#	0.11	32	^-5	k_v	0.28	51	m-n	0_D	0.14
7	h-.	i_2	0.07	32	0-9	v_r	0.29	51	D-d	m_0	0.05
8	d-.	n_#	0.09	32	.-t	r_#	0.07	51	&-E	m_n	0.2
9	j-.	p_u	0.09	33	p-k	#_9	0.23	52	t-d	r_#	0.09
9	u-8	j_r	0.29	34	s-k	#_9	0.27	53	E-a	h_l	0.27
10	E-&	v_s	0.19	35	D-d	#_I	0.05	54	O-0	p_r	0.25
11	.-d	b_a	0.11	36	E-I	d_s	0.2	54	r-.	O_?	0.04
11	0-.	n_b	0.07	36	s-z	E_I	0.07	54	?-t	r_0	0.17
11	b-.	0_l	0.21	36	I-.	s_m	0.11	55	E-^	s_p	0.26
12	1-0	#_s	0.2	37	.-d	#_2	0.11	55	p-f	E_0	0.23
12	i-I	s_d	0.15	38	&-E	#_l	0.2	56	.-I	#_.	0.1
13	j-.	k_u	0.09	39	&-a	r_?	0.18	56	.-n	._v	0.1
14	d-t	n_#	0.11	40	a-5	r_J	0.27	57	&-u	J_v	0.3
15	1-0	#_l	0.2	40	J-.	a_0	0.52	57	.-0	v_l	0.04
16	E-&	#_l	0.19	40	0-.	J_n	0.07	57	l-n	._I	0.23
17	.-r	0_t	0.03	40	.-i	n_0	0.13	57	I-1	l_n	0.28
17	z-Z	i_0	0.38	41	I-i	#_n	0.1	57	n-l	I_#	0.31
18	h-.	#_O	0.07	41	n-.	I_t	0.1	58	t-.	E_s	0.07
18	O-0	w_r	0.25	42	I-E	#_f	0.21	58	s-z	t_b	0.07
19	O-0	w_r	0.25	42	f-v	I_0	0.17	59	h-m	#_&	0.38
20	O-5	k_z	0.26	42	d-n	0_I	0.38	60	.-m	#_i	0.14
20	z-s	O_#	0.07	42	I-E	d_z	0.21	61	I-E	m_n	0.21
21	v-b	#_E	0.18	42	z-s	I_0	0.07	62	.-m	#_.	0.14
21	r-l	E_I	0.11	43	h-.	#_i	0.07	62	.-0	._r	0.04
21	l-.	k_5	0.08	43	w-v	i_I	0.08	62	z-s	5_#	0.07
22	i-I	b_n	0.15	44	E-&	f_l	0.19	63	v-z	4_0	0.52
23	z-s	3_t	0.07	45	s-f	#_E	0.27	64	O-&	n_?	0.27
23	.-r	0_0	0.03	45	E-i	s_k	0.28	65	.-k	I_.	0.1
24	j-b	#_a	0.66	46	.-l	f_7	0.06	65	.-0	._.	0.04
25	.-r	b_i	0.03	47	.-0	n_.	0.04	65	.-l	._#	0.06

66	a-i	n_n	0.29
66	n-.	a_#	0.1
67	m-n	#_I	0.14
68	d-n	#_^	0.38
69	5-a	#_l	0.25
69	l-.	5_d	0.08
70	r-0	p_0	1
70	0-.	r_t	1
71	0-O	#_r	0.27
71	4-I	r_n	0.28
72	a-&	p_s	0.23
73	1-0	n_l	0.2
74	.-p	#_.	0.16
74	.-l	._.	0.06
74	.-0	._t	0.04
75	.-p	#_.	0.16
75	.-0	._l	0.04
75	?-t	l_I	0.17
75	I-.	?_d	0.11
75	d-.	I_#	0.09
76	.-t	s_0	0.07
76	k-t	0_j	0.16
77	.-j	p_^	0.1
77	^-u	._b	0.28
77	l-.	b_I	0.08
78	k-p	#_j	0.24
79	.-k	#_.	0.1
79	.-w	._4	0.11
80	s-k	#_.	0.27
80	.-w	s_I	0.11
81	.-k	#_w	0.1
82	p-r	#_7	0.66
83	d-r	#_I	0.44
84	n-m	1_d	0.11
85	E-I	r_d	0.2
86	i-1	r_l	0.3
87	5-u	r_d	0.27
88	I-.	u_n	0.11
89	s-S	#_O	0.09
90	p-.	s_I	0.24
90	.-I	n_s	0.1
91	S-s	#_.	0.17
91	.-n	S_4	0.1
92	0-.	#_s	0.07
92	t-d	r_I	0.09
93	.-s	#_2	0.08
94	0-.	S_n	0.07
94	.-7	n_.	1
94	.-r	._.	0.03
94	.-I	._#	0.1
95	T-t	#_9	0.18
95	z-s	r_#	0.07
96	E-I	t_n	0.2
97	T-t	#_O	0.18
97	t-r	O_S	0.34
97	S-C	t_i	0.19
97	i-u	S_w	0.3
97	w-.	i_0	0.13
97	z-s	0_#	0.07
98	2-8	t_r	0.3
99	.-r	t_a	0.03
100	b-.	#_^	0.21
101	r-v	#_&	0.44
102	w-v	#_E	0.08
102	E-&	w_s	0.19
103	d-t	s_#	0.11
104	w-v	#_1	0.08
104	I-i	r_0	0.1
105	w-v	#_I	0.08
105	z-s	0_#	0.07
106	l-v	#_5	0.44
107	a-^	v_l	0.26
107	.-r	0_#	0.03
108	d-v	#_1	0.44
109	9-i	w_r	0.3
110	s-r	#_p	0.66
110	p-.	s_&	0.24
112	O-a	n_?	0.24
112	?-t	O_I	0.17
112	I-i	?_#	0.1
113	.-l	5_.	0.06
113	.-d	._b	0.11
113	5-3	b_z	0.29
114	.-i	d_0	0.13
114	.-n	0_p	0.1
115	t-k	s_5	0.14
116	5-&	p_s	0.29
116	0-&	m_n	0.23
117	.-0	k_r	0.04
117	I-i	t_#	0.1
118	s-S	r_0	0.09
118	0-i	s_r	0.3
119	1-0	#_m	0.2
119	.-i	b_0	0.13
119	N-n	0_#	0.14
120	.-r	a_D	0.03
121	f-.	I_T	0.44
121	T-t	f_#	0.18
122	D-j	n_0	1
122	0-u	D_w	0.26
122	I-E	w_n	0.21
123	.-n	0_t	0.1
124	0-I	1_n	0.01
124	n-N	0_#	0.19
125	n-r	#_1	0.31
126	t-d	&_#	0.09
127	0-I	l_n	0.01
127	n-.	0_t	0.1
128	r-w	#_2	0.14
130	4-1	m_0	0.27
130	0-.	4_n	0.07
130	4-1	n_z	0.27
131	.-l	b_1	0.06
132	l-d	#_I	0.44
133	.-E	#_s	0.38
134	.-r	a_#	0.03
135	I-i	#_C	0.1
136	r-.	a_m	0.04
136	r-.	0_#	0.04
137	r-l	0_u	0.11
138	p-f	m_#	0.23
139	4-0	#_?	0.28
139	?-t	4_I	0.17
139	I-i	?_d	0.1
140	a-5	g_s	0.27
141	^-a	w_n	0.21
142	h-.	t_5	0.07
143	a-u	g_n	0.29
144	4-3	r_d	0.29
145	2-u	b_t	0.29
146	l-n	0_5	0.23
147	1-0	#_w	0.2
147	w-v	1_a	0.08
147	a-O	w_n	0.25
148	g-b	#_0	0.44

148	m-g	0_O	0.66
149	h-w	#_2	0.38
150	t-p	#_I	0.18
151	d-.	#_r	0.09
152	T-j	#_&	1
153	h-j	#_E	0.44
154	T-z	#_I	0.66
155	d-b	&_0	0.21
156	0-4	#_k	0.28
156	9-0	k_r	0.29
157	m-.	#_&	0.19
157	t-k	&_r	0.14
158	.-k	&_l	0.1
158	l-r	._I	0.13
159	h-.	#_&	0.07
160	h-.	#_E	0.07
160	E-4	h_l	0.26
161	0-9	t_r	0.29
161	4-0	n_t	0.28
162	1-&	#_m	0.22
162	&-0	n_s	0.24
163	0-5	t_#	0.23
164	r-.	a_s	0.04
165	O-a	#_r	0.24
165	0-.	r_g	0.07
166	h-.	#_O	0.07
166	O-a	h_r	0.24
167	k-p	s_I	0.24
167	I-0	k_n	0
168	g-.	#_r	0.38
168	r-.	g_&	0.04
169	E-&	#_k	0.19
170	.-r	0_#	0.03
171	4-a	n_J	0.29
171	J-Z	4_#	0.31
172	^-0	g_l	0.27
173	I-0	b_n	0
173	E-&	n_n	0.19
173	I-0	n_#	0
174	9-a	b_r	0.29
175	4-0	g_n	0.28
176	&-E	b_k	0.2
177	I-i	g_n	0.1
178	.-0	b_r	0.04
179	9-E	b_r	0.29
180	.-t	I_#	0.07
181	1-0	m_.	0.2
181	.-r	1_#	0.03
182	O-5	b_r	0.26
182	r-.	O_n	0.04
183	h-.	d_E	0.07
184	a-9	b_r	0.29
185	T-n	r_I	1
186	a-&	k_r	0.23
186	O-0	r_t	0.25
187	a-O	k_r	0.25
187	r-.	a_k	0.04
188	h-.	t_O	0.07
189	i-7	C_r	0.29
190	a-7	C_r	0.29
190	I-0	m_N	0
190	N-n	I_#	0.14
191	C-S	#_&	0.15
191	?-d	n_0	0.12
191	k-.	0_l	0.18
192	.-l	0_r	0.06
193	k-C	#_a	0.38
194	S-C	#_5	0.19
194	5-2	S_d	0.29
195	S-C	#_^	0.19
195	v-b	^_#	0.18
196	^-a	k_m	0.21
197	d-v	n_0	0.44
197	n-r	0_s	0.31
198	h-k	#_2	0.38
199	.-r	k_&	0.03
200	l-r	k_&	0.13
201	.-r	k_i	0.03
202	i-I	r_m	0.15
202	I-0	m_n	0
202	O-0	n_l	0.25
203	O-^	k_r	0.28
203	r-.	O_J	0.04
204	p-b	m_0	0.14
205	a-&	d_l	0.23
205	r-.	0_z	0.04
205	z-s	r_#	0.07
206	^-&	d_n	0.27
207	b-d	#_i	0.38
208	w-d	#_E	1
209	r-d	#_I	0.29
210	D-d	#_I	0.05
210	s-S	I_I	0.09
211	1-&	v_n	0.22
212	0-&	m_n	0.23
212	n-.	0_t	0.1
213	.-r	d_u	0.03
213	t-d	I_#	0.09
214	I-0	n_T	0
215	t-k	s_#	0.14
216	D-d	#_0	0.05
216	0-I	D_b	0.01
216	U-0	b_k	0.28
217	k-g	i_w	0.1
217	w-.	k_0	0.13
218	h-.	#_6	0.07
219	w-.	#_i	0.13
220	j-.	#_u	0.09
220	u-i	j_s	0.3
220	.-r	0_#	0.03
221	I-E	#_t	0.21
221	t-.	I_s	0.07
221	s-S	t_0	0.09
221	N-n	a_#	0.14
222	k-t	s_0	0.16
222	0-I	k_.	0.01
222	.-k	0_#	0.1
223	0-i	#_j	0.24
223	j-.	0_O	0.09
223	O-0	j_n	0.25
224	9-E	d_r	0.29
224	r-.	9_m	0.04
225	g-t	0_r	0.66
225	r-.	g_&	0.04
225	m-f	&_#	0.44
226	J-g	r_#	0.66
227	.-0	k_l	0.04
228	I-E	#_f	0.21
229	.-f	#_4	0.27
230	a-&	f_r	0.23
230	r-.	a_m	0.04
230	N-n	I_#	0.14
231	I-1	l_n	0.28
232	I-.	n_S	0.11
232	S-C	I_#	0.19

233	T-f	#_I	0.24
234	.-j	f_O	0.1
235	r-l	f_E	0.11
236	9-5	l_r	0.3
236	r-.	9_t	0.04
237	0-.	f_r	0.07
237	r-.	0_5	0.04
238	2-a	f_n	0.27
238	d-t	n_#	0.11
239	d-.	n_z	0.09
239	i-I	z_#	0.15
240	r-.	g_2	0.04
241	0-1	t_l	0.29
242	t-b	a_#	0.44
243	a-I	g_?	0.28
243	?-t	a_0	0.17
243	0-a	?_.	0.24
243	.-r	0_#	0.03
244	z-C	l_#	1
245	f-p	l_#	0.21
246	E-4	h_l	0.26
247	E-4	h_l	0.26
248	.-h	#_&	0.08
249	.-h	#_a	0.08
249	a-&	._m	0.23
250	.-r	a_m	0.03
251	t-?	r_.	0.21
251	.-0	t_b	0.04
251	.-s	i_t	0.08
252	.-l	E_m	0.06
253	k-h	#_O	0.34
254	2-O	h_.	0.3
254	.-r	2_s	0.03
255	I-i	1_n	0.1
256	4-1	#_k	0.27
256	O-a	k_r	0.24
256	r-.	O_n	0.04
257	h-.	#_1	0.07
258	1-I	#_m	0.28
258	m-g	1_g	0.66
258	g-w	m_a	1
259	I-E	s_s	0.21
260	d-t	a_#	0.11
261	E-0	k_t	0.23
262	C-J	#_i	0.18
263	^-E	J_s	0.29
264	r-J	#_I	0.38
264	k-g	I_S	0.1
264	S-s	k_O	0.17
265	^-a	J_g	0.21
266	z-s	u_#	0.07
267	I-.	l_p	0.11
267	r-.	0_#	0.04
268	.-0	g_r	0.04
269	u-I	b_t	0.29
270	&-I	k_n	0.28
271	s-S	r_#	0.09
272	&-I	k_C	0.28
273	E-&	l_?	0.19
274	r-.	0_#	0.04
275	C-J	E_0	0.18
276	s-k	0_I	0.27
276	I-O	s_.	0.3
276	.-n	I_#	0.1
277	t-.	E_s	0.07
278	U-I	l_k	0.29
278	k-g	U_w	0.1
278	w-.	k_0	0.13
279	2-0	m_s	0.29
280	w-l	#_I	0.44
281	9-0	k_r	0.29
281	r-.	9_s	0.04
281	.-t	s_#	0.07
282	s-l	#_a	0.66
283	.-n	g_0	0.1
284	.-m	#_&	0.14
285	r-l	0_I	0.11
286	^-a	w_n	0.21
286	.-0	n_#	0.04
287	^-a	m_.	0.21
287	.-r	^_m	0.03
287	O-0	m_t	0.25
288	^-&	m_s	0.27
288	r-.	0_d	0.04
289	d-m	#_4	0.66
290	s-z	I_0	0.07
291	0-5	m_m	0.23
291	5-0	?_.	0.25
291	.-m	5_#	0.14
292	.-0	n_s	0.04
292	0-7	t_r	0.29
293	&-a	m_n	0.18
293	n-N	&_g	0.19
294	^-u	m_s	0.28
295	?-t	r_I	0.17
295	I-a	?_.	0.28
295	.-r	I_f	0.03
296	^-a	m_s	0.21
296	s-T	^_#	0.11
297	.-l	^_C	0.06
298	w-z	^_I	1
298	I-0	w_l	0
299	I-.	u_t	0.11
300	&-I	n_n	0.28
301	5-a	#_z	0.25
301	z-s	5_0	0.07
301	a-0	l_t	0.26
302	1-3	#_l	0.28
303	&-a	#_m	0.18
303	j-.	m_u	0.09
303	u-.	j_l	1
304	d-n	a_0	0.38
304	w-m	0_a	1
304	a-&	w_n	0.23
304	n-?	a_0	1
304	.-0	i_#	0.04
305	h-.	#_5	0.07
305	1-0	p_l	0.2
306	p-.	#_O	0.24
306	g-J	r_I	0.44
307	E-0	v_r	0.23
309	4-I	r_J	0.28
310	I-0	t_l	0
311	s-T	I_#	0.11
312	4-1	l_0	0.27
313	f-p	#_0	0.21
313	0-a	f_r	0.24
313	r-.	0_s	0.04
313	^-0	s_m	0.27
314	^-a	s_m	0.21
314	b-.	m_0	0.21
315	z-S	I_m	0.31
315	E-0	m_n	0.23
316	.-0	1_r	0.04
316	s-.	0_#	0.1

317	.-w	k_a	0.11
318	.-k	#_w	0.1
318	^-a	w_n	0.21
318	&-0	d_m	0.24
319	h-.	r_&	0.07
319	&-4	h_p	0.27
320	I-i	r_p	0.1
321	k-g	N_0	0.1
322	I-i	r_f	0.1
322	9-0	f_r	0.29
323	d-l	i_I	0.34
324	.-j	f_0	0.1
324	0-U	._r	0.29
324	r-J	0_i	0.38
325	v-b	I_0	0.18
325	.-d	l_r	0.11
326	.-0	k_S	0.04
326	O-4	S_#	0.29
327	S-s	^_t	0.17
328	5-0	s_f	0.25
328	.-I	r_#	0.1
329	^-a	s_m	0.21
330	.-s	#_k	0.08
330	.-r	k_O	0.03
331	t-k	s_r	0.14
332	i-0	d_m	0.26
332	E-0	m_n	0.23
333	4-E	s_m	0.28
333	b-.	m_0	0.21
334	t-n	0_7	0.34
335	.-0	n_t	0.04
335	O-0	t_r	0.25
336	i-.	S_&	0.29
337	i-I	S_E	0.15
337	E-.	i_v	1
338	s-S	#_5	0.09
338	a-0	f_r	0.26
339	.-t	r_#	0.07
340	a-O	S_.	0.25
340	.-r	a_t	0.03
341	5-a	S_.	0.25
341	.-t	5_g	0.07
342	.-I	s_m	0.1
342	1-0	m_l	0.2
342	.-i	l_#	0.13
343	d-t	l_#	0.11
344	D-d	5_0	0.05
345	O-a	s_D	0.24
345	D-d	O_0	0.05
346	.-w	s_a	0.11
346	I-4	r_#	0.28
347	E-a	s_l	0.27
347	l-.	E_f	0.08
348	.-s	#_.	0.08
348	.-p	._E	0.16
348	s-.	E_k	0.1
348	.-j	k_0	0.1
349	5-9	t_.	0.29
349	.-r	5_n	0.03
351	w-.	m_^	0.13
351	^-0	w_n	0.27
352	0-i	t_b	0.24
353	d-t	n_.	0.11
353	.-r	d_0	0.03
354	D-t	#_7	0.18
354	7-9	D_r	0.29
355	I-.	5_t	0.11
356	T-t	#_I	0.18
357	r-.	a_g	0.04
358	^-a	t_g	0.21
359	d-t	#_a	0.11
360	^-a	t_n	0.21
360	u-0	S_r	0.3
361	O-0	1_l	0.25
362	.-t	#_.	0.07
362	.-r	._2	0.03
363	J-C	r_I	0.2
364	u-U	j_r	0.29
365	w-v	#_i	0.08
366	w-v	#_E	0.08
366	E-&	w_n	0.19
367	m-v	#_0	1
368	l-v	#_&	0.44
369	f-v	#_i	0.17
370	.-v	#_.	0.34
370	.-0	._.	0.04
370	.-r	._m	0.03
371	J-Z	r_0	0.31
372	w-v	#_1	0.08
373	w-v	#_a	0.08
373	t-d	a_k	0.09
374	t-.	4_f	0.07
375	h-w	#_I	0.38
376	E-i	w_n	0.28
377	n-w	#_I	0.52
378	E-I	w_n	0.2
379	I-E	r_C	0.21
379	0-I	C_r	0.01
379	r-.	0_d	0.04
380	?-D	1_I	1
381	h-j	#_i	0.44
381	i-.	h_O	0.29
381	O-a	i_t	0.24
382	u-a	j_r	0.29
382	a-.	r_n	0.66
383	h-j	#_E	0.44
384	d-k	5_0	0.34
385	s-z	#_E	0.07
386	s-z	#_I	0.07
387	s-z	#_^	0.07
387	^-a	s_m	0.21
388	s-z	#_2	0.07
390	0-O	#_l	0.27
390	.-r	0_#	0.03
404	k-t	#_r	0.16
405	s-.	b_0	0.1
406	.-d	&_m	0.11
406	j-.	m_u	0.09
406	u-0	j_n	0.28
407	I-0	#_l	0
408	a-5	m_n	0.27
408	E-0	j_t	0.23
408	t-.	E_#	0.07
409	0-&	#_n	0.23
410	.-k	s_E	0.1
411	s-.	#_0	0.1
411	v-.	r_1	0.17
412	.-p	s_#	0.16
413	s-b	0_u	0.66
414	.-r	0_E	0.03
414	E-I	._t	0.2
415	^-O	b_b	0.28
416	d-b	#_4	0.21
417	p-b	#_i	0.14
418	N-n	i_z	0.14

419	a-i	b_r	0.3
420	^-I	b_g	0.27
421	1-a	l_n	0.23
422	E-4	r_d	0.26
423	0-a	r_z	0.24
423	z-Z	0_6	0.38
423	i-0	z_r	0.29
424	d-.	E_T	0.09
425	d-b	#_r	0.21
426	s-.	E_t	0.1
426	t-T	s_s	0.2
427	k-g	5_0	0.1
428	l-r	b_u	0.13
429	n-N	^_d	0.19
429	d-g	n_0	0.44
430	I-U	b_C	0.3
431	m-k	#_&	0.66
432	t-p	r_#	0.18
433	p-m	I_i	0.66
433	d-?	i_.	0.18
433	.-0	d_.	0.04
433	.-r	._#	0.03
434	.-m	I_p	0.14
435	&-I	C_n	0.28
435	.-t	n_s	0.07
436	r-l	k_2	0.11
437	d-.	l_#	0.09
438	d-t	l_#	0.11
439	m-n	a_p	0.14
439	p-d	m_l	1
439	l-.	p_I	0.08
440	.-0	k_.	0.04
440	.-n	._w	0.1
440	w-J	._E	0.66
441	h-k	#_5	0.38
442	C-k	#_a	0.52
442	a-O	C_r	0.25
442	g-s	r_0	1
442	0-I	g_z	0.01
442	z-t	0_#	1
443	s-t	r_a	0.44
443	a-E	s_Z	0.29
444	g-k	#_r	0.11
445	k-?	I_I	0.66
445	t-k	I_#	0.14
446	C-k	#_.	0.52
446	.-r	C_u	0.03
447	O-^	r_l	0.28
448	d-b	^_#	0.21
449	J-z	I_#	1
450	O-9	k_r	0.29
451	D-d	#_0	0.05
451	0-I	D_l	0.01
452	D-d	#_0	0.05
452	0-I	D_p	0.01
453	E-I	d_n	0.2
454	I-1	d_n	0.28
454	r-.	0_s	0.04
455	u-5	d_#	0.27
456	l-d	#_a	0.44
457	t-d	#_2	0.09
458	.-r	d_&	0.03
459	O-u	r_l	0.29
460	n-N	^_#	0.19
461	3-9	#_.	0.3
461	.-r	3_l	0.03
462	j-.	#_6	0.09
463	h-.	#_i	0.07
464	r-l	I_E	0.11
465	k-f	0_w	0.44
465	w-.	k_0	0.13
466	&-E	#_m	0.2
467	0-i	#_p	0.24
467	t-.	s_0	0.07
468	I-E	#_C	0.21
469	D-T	i_0	0.23
470	.-r	6_#	0.03
471	p-f	#_a	0.23
472	&-O	f_s	0.29
473	D-v	r_0	0.38
474	w-f	#_1	0.34
475	.-n	I_S	0.1
475	S-C	._#	0.19
476	g-S	I_j	1
476	j-.	g_0	0.09
477	v-f	#_.	0.14
477	.-l	v_1	0.06
478	h-f	#_5	0.52
479	4-3	f_b	0.29
480	T-t	r_#	0.18
481	O-2	f_l	0.29
482	u-2	f_l	0.29
483	E-a	r_n	0.27
484	.-f	#_j	0.27
485	.-m	4_#	0.14
486	0-3	g_l	0.3
487	f-g	#_u	0.52
488	d-g	#_.	0.44
488	.-l	d_&	0.06
489	.-g	#_r	0.18
489	r-l	._u	0.11
490	r-.	a_d	0.04
491	w-g	#_9	0.66
491	9-O	w_r	0.29
492	4-1	r_p	0.27
493	k-g	#_.	0.1
493	.-r	k_3	0.03
494	5-2	r_s	0.29
495	k-h	#_4	0.34
496	4-&	h_s	0.27
496	s-S	4_t	0.09
496	t-.	s_#	0.07
497	S-h	#_i	0.66
498	7-E	h_r	0.28
498	1-I	r_s	0.28
498	i-I	s_#	0.15
499	l-h	#_I	1
500	5-O	h_.	0.28
500	.-r	5_z	0.03
500	z-s	._I	0.07
501	E-^	h_l	0.26
502	n-r	0_#	0.31
503	I-9	h_.	0.3
503	.-r	I_?	0.03
504	b-.	#_1	0.21
505	m-.	#_I	0.19
506	.-I	#_.	0.1
506	.-m	._p	0.14
507	.-r	0_d	0.03
507	.-t	k_#	0.07
508	.-I	#_.	0.1
508	.-n	._v	0.1
509	w-.	#_9	0.13
510	h-.	k_E	0.07
510	E-0	h_l	0.23

511	r-J	#_&	0.38
512	t-J	#_I	1
513	b-J	#_^	1
514	d-t	l_#	0.11
515	s-.	#_k	0.1
516	^-0	d_m	0.27
517	p-k	#_I	0.23
518	&-E	J_s	0.2
519	&-I	l_m	0.28
520	.-d	n_0	0.11
521	t-p	s_#	0.18
522	l-.	0_m	0.08
523	E-a	l_J	0.27
524	m-l	#_u	0.38
525	a-^	l_s	0.26
526	t-k	I_4	0.14
527	h-.	n_u	0.07
527	u-0	h_p	0.28
528	1-I	m_l	0.28
529	&-I	m_r	0.28
529	I-a	r_J	0.28
529	J-Z	I_#	0.31
530	I-.	5_n	0.11
531	.-t	5_s	0.07
532	&-a	m_n	0.18
533	.-k	s_#	0.1
534	I-^	m_s	0.29
535	.-r	a_k	0.03
536	z-s	i_#	0.07
537	l-n	#_1	0.23
538	t-d	5_#	0.09
539	w-n	#_^	0.52
540	h-n	#_O	0.52
541	0-&	v_n	0.23
542	7-&	p_r	0.26
542	t-k	0_I	0.14
542	I-i	t_t	0.1
543	.-t	0_s	0.07
544	n-.	0_t	0.1
545	^-O	p_n	0.28
546	5-E	p_s	0.29
547	.-r	f_4	0.03
548	n-.	0_t	0.1
548	i-I	t_#	0.15
549	w-p	#_I	1
550	f-p	#_u	0.21
551	0-u	p_r	0.26
551	I-i	r_#	0.1
552	a-E	r_p	0.29
553	n-m	0_#	0.11
554	g-p	#_^	0.66
555	b-p	#_^	0.16
556	.-r	1_.	0.03
556	.-5	._m	0.24
557	.-w	k_4	0.11
558	b-k	#_.	0.52
558	.-w	b_&	0.11
558	&-O	._r	0.29
559	.-k	#_w	0.1
560	i-7	w_r	0.29
560	.-I	r_#	0.1
561	.-k	#_w	0.1
562	.-k	#_w	0.1
563	g-k	#_.	0.11
563	.-w	g_I	0.11
564	h-r	#_&	0.66
565	.-r	#_.	0.03
565	.-&	._.	0.52
565	.-k	._m	0.1
565	^-O	n_f	0.28
566	T-f	i_#	0.24
567	v-z	E_0	0.52
568	p-b	I_a	0.14
569	5-a	b_t	0.25
570	b-r	#_U	0.66
571	t-.	#_r	0.07
572	&-^	r_p	0.28
573	5-u	r_z	0.27
574	I-^	r_s	0.29
575	&-E	#_s	0.2
576	.-I	l_v	0.1
577	.-0	s_r	0.04
578	t-k	s_a	0.14
579	.-s	#_h	0.08
579	h-k	._a	0.38
579	0-I	l_t	0.01
580	S-s	#_.	0.17
580	.-k	S_9	0.1
580	9-7	._r	0.3
581	S-s	#_.	0.17
581	.-k	S_r	0.1
582	u-a	r_d	0.3
583	C-S	#_E	0.15
584	a-5	S_t	0.27
585	J-s	#_I	1
586	T-s	#_I	0.25
587	0-a	?_r	0.24
588	S-s	#_.	0.17
588	.-n	S_4	0.1
589	7-O	n_r	0.29
590	.-s	#_.	0.08
590	.-p	._l	0.16
591	a-O	p_.	0.25
591	.-r	a_t	0.03
592	.-s	#_k	0.08
592	.-w	k_I	0.11
593	^-a	t_k	0.21
594	.-s	#_w	0.08
595	O-&	t_.	0.27
595	.-l	O_k	0.06
596	m-t	#_&	1
597	p-t	#_4	0.31
598	1-i	t_m	0.28
599	0-.	t_i	0.07
600	s-z	n_#	0.07
601	I-O	r_l	0.3
602	r-t	#_1	0.52
603	n-t	#_1	0.66
603	n-m	1_#	0.11
604	n-m	0_#	0.11
605	b-t	#_^	0.38
606	k-f	1_u	0.44
607	r-.	#_^	0.04
607	.-r	0_#	0.03
608	O-u	j_r	0.29
609	.-t	I_#	0.07
610	b-v	#_4	0.44
611	b-v	#_1	0.44
611	.-t	k_#	0.07
612	s-.	#_p	0.1
612	p-v	s_1	0.66
613	0-.	1_l	0.07
613	I-.	l_n	0.11
613	n-.	I_z	0.1
614	w-v	#_I	0.08

615	f-v	#_a	0.17
616	f-v	#_O	0.17
616	O-5	f_l	0.26
617	9-O	w_r	0.29
618	t-h	#_8	1
618	u-O	t_r	0.29
619	l-w	#_1	0.25
620	v-w	#_1	1
621	w-r	#_I	0.21
622	s-T	u_#	0.11
623	s-z	#_i	0.07
623	i-I	s_n	0.15
631	.-s	#_t	0.08
632	0-E	t_r	0.26
632	r-l	0_#	0.11
633	.-h	0_E	0.08
633	v-f	E_0	0.14
634	h-.	p_I	0.07
635	h-.	#_I	0.07
635	I-&	h_m	0.26
636	.-O	g_r	0.52
636	I-0	r_#	0
637	^-a	b_.	0.21
637	.-r	^_?	0.03
638	.-l	b_&	0.06
639	I-1	d_l	0.28
639	1-4	l_?	0.26
640	.-d	#_r	0.11
641	&-^	r_g	0.28
642	n-.	#_E	0.1
642	s-m	E_0	1
642	0-I	s_s	0.01
643	u-3	f_l	0.3
644	T-f	#_r	0.24
645	.-g	#_&	0.18
646	r-.	0_z	0.04
646	z-s	r_#	0.07
647	k-.	#_0	0.18
647	0-I	k_n	0.01
648	g-t	I_4	0.66
649	.-z	E_b	0.25
649	.-i	b_0	0.13
649	i-.	n_z	0.29
649	z-.	i_#	0.21
650	O-5	#_f	0.26
650	f-v	O_0	0.17
651	r-w	k_4	0.14
651	I-1	z_#	0.28
652	.-k	#_.	0.1
652	.-w	._i	0.11
653	d-k	#_.	0.34
653	.-w	d_6	0.11
654	5-u	r_m	0.27
654	n-.	&_?	0.1
655	h-t	#_a	0.44
656	s-S	n_E	0.09
657	z-D	5_d	0.44
658	E-I	m_n	0.2
659	E-^	m_s	0.26
660	b-p	#_O	0.16
661	^-&	r_f	0.27
662	4-i	r_p	0.27
663	l-n	E_0	0.23
663	?-d	4_I	0.12
664	^-E	S_t	0.29
664	t-d	^_#	0.09
665	.-J	n_d	0.34
666	^-O	f_.	0.28
666	.-r	^_k	0.03
667	I-i	k_l	0.1
668	.-0	#_k	0.04
669	d-.	#_I	0.09
669	I-0	d_n	0
669	n-b	I_a	0.52
670	E-4	#_g	0.26
670	g-k	E_z	0.11
670	z-s	g_#	0.07
671	E-O	#_l	0.3
672	E-&	#_l	0.19
672	0-I	l_v	0.01
672	v-g	0_4	0.66
673	&-4	#_n	0.27
674	&-4	#_N	0.27
674	N-n	&_g	0.14
674	g-J	N_0	0.44
675	l-.	#_&	0.08
676	m-.	#_&	0.19
677	n-b	#_i	0.52
678	r-l	b_E	0.11
679	p-b	#_l	0.14
679	l-r	p_4	0.13
680	J-b	#_^	0.66
681	U-^	b_S	0.28
681	S-s	U_#	0.17
682	r-b	#_1	0.66
683	h-.	S_u	0.07
684	l-.	O_z	0.08
685	v-.	l_z	0.17
686	7-a	C_r	0.29
687	k-C	#_4	0.38
687	4-&	k_.	0.27
687	.-s	4_a	0.08
687	a-i	._s	0.29
687	s-.	a_#	0.1
688	m-C	#_&	1
689	.-l	k_I	0.06
690	g-k	#_5	0.11
691	5-a	k_m	0.25
692	.-k	#_I	0.1
692	I-0	._n	0
693	z-s	I_#	0.07
694	.-r	O_z	0.03
695	O-a	k_r	0.24
695	.-0	r_n	0.04
696	t-k	#_.	0.14
696	.-r	t_1	0.03
697	.-r	k_O	0.03
697	O-^	._l	0.28
697	z-.	r_#	0.21
698	.-k	#_j	0.1
698	.-i	n_0	0.13
699	9-1	s_r	0.29
699	r-.	9_k	0.04
700	.-d	#_I	0.11
701	.-d	#_I	0.11
701	I-i	._n	0.1
701	n-.	I_v	0.1
701	v-t	n_E	0.66
702	.-d	#_2	0.11
703	t-d	#_^	0.09
704	w-.	#_i	0.13
705	I-E	#_l	0.21
706	h-.	#_E	0.07
707	i-E	#_l	0.29
707	E-0	l_k	0.23

707	t-.	k_r	0.07
707	r-.	t_0	0.04
707	0-.	r_k	0.07
707	k-.	0_j	0.18
708	m-.	#_E	0.19
709	.-i	#_m	0.13
710	0-E	#_h	0.26
710	h-n	0_&	0.52
711	b-.	#_1	0.21
711	1-E	b_.	0.28
711	.-n	1_s	0.1
711	.-l	k_.	0.06
711	.-I	._.	0.1
711	.-k	._0	0.1
712	i-E	#_r	0.29
712	.-r	0_#	0.03
713	s-.	#_0	0.1
713	0-E	s_k	0.26
714	t-d	4_#	0.09
715	O-a	f_r	0.24
716	.-n	E_D	0.1
716	D-d	._0	0.05
717	D-?	E_0	0.23
718	.-l	f_&	0.06
719	9-a	f_r	0.29
719	r-.	9_D	0.04
719	D-d	r_0	0.05
720	J-f	#_.	1
720	.-j	J_U	0.1
721	9-&	g_r	0.29
721	r-.	9_l	0.04
722	&-E	J_m	0.2
723	I-.	n_t	0.11
723	t-?	I_0	0.21
723	0-1	t_l	0.29
724	r-g	#_I	0.52
725	b-g	#_l	0.31
726	.-r	g_4	0.03
727	k-g	#_l	0.1
727	l-r	k_&	0.13
728	k-g	#_r	0.1
729	t-d	l_#	0.09
730	f-h	#_7	0.38
731	.-m	l_#	0.14
732	2-5	h_s	0.29
732	s-z	2_#	0.07
733	n-h	#_5	0.52
734	d-l	I_0	0.34
735	r-s	0_#	0.44
736	&-^	h_N	0.28
737	l-.	f_0	0.08
737	.-r	0_m	0.03
738	C-J	#_E	0.18
739	l-J	#_I	1
740	1-a	n_t	0.23
741	v-b	&_.	0.18
741	.-r	v_0	0.03
741	t-d	0_O	0.09
741	I-.	r_z	0.11
742	b-.	#_r	0.21
742	r-l	b_i	0.11
743	v-n	I_I	1
743	I-0	v_N	0
743	N-n	I_#	0.14
744	w-l	#_u	0.44
745	.-r	O_#	0.03
746	n-m	#_4	0.11
746	m-n	4_#	0.14
747	z-s	n_#	0.07
748	&-i	m_?	0.3
749	.-0	n_.	0.04
749	.-r	._I	0.03
749	I-E	._t	0.21
750	n-l	a_0	0.31
750	0-I	n_k	0.01
750	.-j	k_0	0.1
750	0-u	._l	0.26
751	^-5	m_n	0.28
751	I-4	n_#	0.28
752	m-n	^_I	0.14
753	0-5	m_n	0.23
753	?-g	a_0	0.66
753	n-m	0_I	0.11
754	E-I	n_s	0.2
754	t-.	s_#	0.07
755	g-n	#_&	1
756	b-n	#_&	0.66
758	.-n	#_1	0.1
758	1-a	._d	0.23
758	d-n	1_E	0.38
759	0-5	#_d	0.23
759	O-0	d_r	0.25
760	.-h	p_0	0.08
760	0-5	._l	0.23
761	r-p	#_I	0.52
761	I-0	r_.	0
761	.-r	I_z	0.03
761	z-s	._I	0.07
762	.-0	i_n	0.04
763	.-0	k_.	0.04
763	.-r	._0	0.03
764	0-5	l_r	0.23
764	r-.	0_#	0.04
765	k-.	I_C	0.18
766	s-p	#_I	1
767	f-p	#_.	0.21
767	.-l	f_4	0.06
768	.-p	#_.	0.16
768	.-l	._&	0.06
768	l-.	&_?	0.08
769	.-l	p_a	0.06
770	&-u	p_d	0.3
771	I-.	5_.	0.11
771	.-l	I_t	0.06
772	z-s	u_#	0.07
773	s-r	#_i	0.66
774	z-s	i_0	0.07
774	n-s	0_#	1
775	C-J	I_I	0.18
775	.-d	I_z	0.11
776	u-5	r_m	0.27
777	E-^	r_n	0.26
778	.-k	s_&	0.1
779	t-k	s_a	0.14
780	p-k	s_.	0.23
780	.-r	p_i	0.03
781	.-n	E_d	0.1
781	d-?	._0	0.18
782	.-s	#_i	0.08
783	a-&	S_k	0.23
784	I-&	v_n	0.26
785	.-r	S_I	0.03
786	.-s	#_r	0.08
786	r-l	._^	0.11
787	n-m	s_&	0.11

788	.-I	4_n	0.1
788	n-N	._#	0.19
789	.-s	#_t	0.08
789	t-p	._r	0.18
790	a-7	t_r	0.29
791	.-s	#_k	0.08
791	k-t	._I	0.16
792	m-k	r_#	0.66
793	E-2	r_s	0.3
794	k-t	s_1	0.16
795	t-d	i_#	0.09
796	E-I	s_l	0.2
797	O-a	t_l	0.24
797	^-U	m_d	0.29
798	d-T	#_r	0.66
798	r-.	d_O	0.04
799	4-1	t_l	0.27
800	m-n	1_#	0.14
801	1-4	r_0	0.26
801	0-.	1_l	0.07
802	.-r	t_E	0.03
802	t-d	n_#	0.09
803	.-v	#_E	0.34
803	E-&	._m	0.19
804	f-v	#_O	0.17
805	w-v	#_I	0.08
805	I-E	w_s	0.21
806	w-v	#_4	0.08
806	4-i	w_1	0.27
807	r-v	#_I	0.44
807	k-t	I_u	0.16
808	p-.	#_l	0.24
808	l-w	p_4	0.25
809	h-w	#_E	0.38
809	m-d	E_l	0.44
810	.-w	#_i	0.11
811	p-k	s_0	0.23
812	a-O	h_r	0.25
812	0-.	r_s	0.07
813	z-f	1_#	1
814	.-r	d_5	0.03
815	l-j	#_9	0.25
816	d-b	r_0	0.21
819	E-U	p_l	0.3
819	.-k	l_0	0.1
819	0-5	._#	0.23
820	h-.	#_i	0.07
820	.-I	i_d	0.1
821	1-&	#_g	0.22
821	a-I	g_t	0.28
822	0-&	#_n	0.23
822	I-0	d_l	0
823	E-0	g_s	0.23
824	1-a	#_.	0.23
824	.-r	1_m	0.03
824	?-d	E_I	0.12
824	I-0	?_n	0
825	0-&	#_t	0.23
825	&-0	l_s	0.24
826	h-.	#_&	0.07
826	&-1	h_v	0.29
826	I-0	v_n	0
827	.-0	4_l	0.04
828	&-a	b_N	0.18
828	D-g	N_.	0.44
828	.-l	D_0	0.06
828	I-E	d_S	0.21
829	.-0	k_r	0.04
830	.-l	p_4	0.06
831	C-S	#_a	0.15
831	a-&	C_m	0.23
831	.-l	p_I	0.06
831	I-4	._N	0.28
831	N-n	I_#	0.14
832	.-i	k_&	0.13
832	&-a	._n	0.18
832	j-.	t_E	0.09
832	E-i	j_#	0.28
833	s-z	E_t	0.07
833	t-d	s_I	0.09
833	I-0	t_n	0
834	&-^	d_n	0.28
835	&-E	#_s	0.2
835	4-0	s_k	0.28
836	a-&	h_n	0.23
836	n-l	a_0	0.31
836	l-j	u_u	0.25
836	u-0	l_#	0.28
837	d-v	r_0	0.44
837	.-d	r_#	0.11
838	E-0	m_n	0.23
838	I-4	n_#	0.28
839	0-5	k_r	0.23
839	r-d	0_a	0.29
839	a-&	r_n	0.23
839	n-k	a_#	0.66
840	U-u	k_d	0.29
840	j-.	d_u	0.09
841	I-a	w_k	0.28
841	k-t	I_#	0.16
842	z-s	&_I	0.07
843	4-E	m_k	0.28
843	r-.	0_#	0.04
844	U-^	t_k	0.28
845	.-5	r_#	0.24
846	4-9	n_.	0.3
846	.-r	4_C	0.03
847	&-a	p_.	0.18
847	.-r	&_s	0.03
847	s-T	._I	0.11
847	I-0	s_n	0
848	k-p	#_a	0.24
848	a-^	k_n	0.26
849	d-.	l_#	0.09
850	s-T	#_I	0.11
851	0-I	t_b	0.01
852	O-U	j_r	0.29
852	.-0	r_.	0.04
852	.-g	._w	0.18
854	9-2	f_r	0.3
854	r-.	9_s	0.04
855	.-h	#_a	0.08
855	w-v	r_0	0.08
856	C-J	#_3	0.18
857	.-0	k_l	0.04
857	u-8	l_#	0.29
858	.-I	#_r	0.1
858	E-a	r_k	0.27
859	u-0	n_.	0.28
859	.-w	u_I	0.11
859	I-0	._n	0
860	&-0	k_r	0.24
860	r-.	0_#	0.04
861	9-O	#_r	0.29
862	t-.	#_r	0.07

863	&-4	l_m	0.27
864	^-&	#_?	0.27
864	?-d	^_0	0.12
865	?-d	&_I	0.12
865	.-t	k_#	0.07
866	r-.	a_m	0.04
867	.-I	b_j	0.1
867	j-v	._0	1
868	t-?	n_i	0.21
868	i-I	t_k	0.15
869	^-O	#_D	0.28
869	D-T	^_0	0.23
870	i-0	b_z	0.26
871	f-b	#_i	0.66
872	v-b	#_E	0.18
872	r-l	E_I	0.11
873	.-b	#_w	0.21
873	w-l	._l	0.44
874	&-a	b_?	0.18
875	.-b	#_9	0.21
875	9-u	._r	0.3
876	D-d	i_0	0.05
876	.-r	0_#	0.03
877	.-k	I_s	0.1
878	g-k	#_l	0.11
879	5-a	k_l	0.25
879	l-.	5_d	0.08
880	&-4	r_S	0.27
881	w-r	s_9	0.21
881	9-5	w_r	0.3
881	r-.	9_d	0.04
882	^-5	r_g	0.28
883	b-d	#_l	0.38
883	l-.	b_u	0.08
884	h-.	#_6	0.07
885	S-.	#_E	0.52
886	.-i	#_p	0.13
887	i-0	n_k	0.26
888	i-E	#_r	0.29
889	k-f	#_&	0.44
890	s-.	#_p	0.1
890	p-f	s_i	0.23
891	b-f	#_I	0.38
892	v-f	#_1	0.14
893	p-f	#_.	0.23

893	.-l	p_i	0.06
894	.-l	f_O	0.06
895	d-.	n_#	0.09
896	.-I	n_?	0.1
896	?-t	._0	0.17
897	r-l	g_&	0.11
898	.-t	f_#	0.07
899	5-^	g_s	0.29
900	h-.	#_7	0.07
901	a-9	h_r	0.29
902	&-I	h_p	0.28
903	f-h	#_&	0.38
903	&-a	f_m	0.18
904	O-2	h_r	0.29
904	r-.	O_s	0.04
905	a-9	h_r	0.29
906	g-r	0_r	1
906	r-v	g_0	0.44
907	h-.	#_1	0.07
908	d-J	#_&	0.66
909	.-J	#_E	0.34
910	O-2	J_.	0.29
910	.-l	O_#	0.06
911	I-^	J_N	0.29
912	E-^	J_t	0.26
913	d-l	#_4	0.34
914	0-I	m_n	0.01
914	n-N	0_#	0.19
915	t-?	n_I	0.21
915	I-0	t_l	0
916	b-l	#_I	0.66
917	l-r	0_#	0.13
918	b-l	#_I	0.66
919	a-0	k_s	0.26
920	5-O	l_0	0.28
920	0-.	5_r	0.07
921	J-l	#_^	1
922	n-N	I_t	0.19
922	t-k	n_#	0.14
923	4-0	n_z	0.28
923	z-s	4_0	0.07
923	0-E	z_l	0.26
924	m-n	#_4	0.14
925	t-p	i_#	0.18
926	E-I	n_t	0.2

927	I-0	d_l	0
928	b-n	#_U	0.66
929	p-n	#_5	0.66
930	f-n	#_^	1
931	l-n	#_I	0.23
932	d-.	l_#	0.09
933	l-.	p_E	0.08
933	s-z	E_0	0.07
934	p-f	#_E	0.23
935	r-l	p_I	0.11
936	u-5	p_r	0.3
936	r-.	8_#	0.04
937	l-w	k_I	0.25
938	p-b	&_I	0.14
938	d-t	I_#	0.11
939	E-&	r_k	0.19
939	0-u	k_n	0.26
940	g-.	#_r	0.38
940	t-d	4_#	0.09
941	i-4	r_l	0.29
942	n-r	#_6	0.31
943	.-r	#_.	0.03
943	.-I	._m	0.1
944	d-b	u_#	0.21
945	t-.	1_#	0.07
946	E-4	s_l	0.26
947	^-&	s_m	0.27
948	t-k	s_i	0.14
949	.-s	#_t	0.08
949	t-k	._.	0.14
949	.-r	t_5	0.03
950	I-i	s_k	0.1
951	s-S	#_&	0.09
952	.-r	a_k	0.03
953	J-S	#_4	0.38
954	p-k	a_#	0.23
955	.-s	#_.	0.08
955	.-I	._.	0.1
955	.-n	._s	0.1
956	S-s	#_.	0.17
956	.-n	S_O	0.1
957	.-j	n_I	0.1
957	I-0	._S	0
957	S-l	I_#	1
958	4-1	p_s	0.27

959	I-i	t_l	0.1
960	a-O	t_r	0.25
960	t-k	r_#	0.14
961	.-w	s_i	0.11
962	S-s	#_.	0.17
962	.-w	S_I	0.11
963	d-.	r_#	0.09
964	b-T	#_1	1
965	t-T	#_r	0.2
966	T-t	#_I	0.18
967	k-t	#_U	0.16
967	U-^	k_N	0.28
968	r-.	a_?	0.04
969	h-t	#_O	0.44
969	O-8	h_r	0.29
970	d-t	#_.	0.11
970	.-r	d_&	0.03
971	s-.	#_t	0.1
972	d-t	#_r	0.11
973	s-t	#_w	0.44
974	l-v	#_&	0.44
975	&-E	v_s	0.2
976	n-v	#_1	0.66
977	D-v	#_i	0.38
977	r-.	0_#	0.04
978	w-v	#_1	0.08
978	z-s	1_#	0.07
979	b-v	#_2	0.44
980	0-.	r_r	0.07
980	r-.	0_#	0.04
981	h-j	#_a	0.44
982	d-j	#_a	1
987	.-d	5_z	0.11
987	z-s	._k	0.07
987	^-a	k_l	0.21
988	&-4	k_n	0.27
988	0-U	C_#	0.29
989	.-h	#_1	0.08
989	^-a	k_p	0.21
990	I-.	1_n	0.11
990	.-r	0_p	0.03
991	0-4	#_C	0.28
991	.-h	C_i	0.08
992	.-0	r_t	0.04
993	I-0	s_J	0
993	0-I	J_n	0.01
994	.-d	i_#	0.11
995	.-m	0_d	0.14
996	.-0	b_r	0.04
997	0-.	k_l	0.07
998	.-l	k_a	0.06
999	g-k	#_.	0.11
999	.-l	g_a	0.06
1000	E-^	b_g	0.26
1001	0-I	p_r	0.01
1001	r-.	0_z	0.04
1002	p-k	i_#	0.23
1003	m-n	i_s	0.14
1004	O-4	#_l	0.29
1005	k-?	r_I	0.66
1005	t-C	I_E	0.38
1005	E-5	t_k	0.29
1005	t-.	k_#	0.07
1006	s-.	#_t	0.1
1006	t-b	s_I	0.44
1007	.-l	b_5	0.06
1008	E-1	r_d	0.3
1009	d-p	r_#	0.44
1010	s-v	n_0	1
1010	v-s	r_4	0.66
1011	.-k	#_r	0.1
1011	.-z	n_#	0.25
1012	p-t	^_#	0.31
1013	z-s	4_#	0.07
1014	D-d	#_0	0.05
1014	j-.	n_0	0.09
1014	0-i	j_.	0.3
1014	.-r	0_#	0.03
1015	f-d	#_a	1
1016	E-i	#_r	0.29
1016	r-.	0_#	0.04
1017	n-f	#_&	0.66
1017	&-9	n_.	0.29
1017	.-r	&_C	0.03
1017	C-t	._0	0.38
1017	0-.	C_r	0.07
1017	r-.	0_0	0.04
1017	0-I	r_l	0.01
1018	.-l	f_4	0.06
1019	r-l	f_u	0.11
1020	E-&	g_s	0.19
1021	b-g	#_r	0.31
1022	l-g	#_^	1
1023	n-m	&_d	0.11
1023	d-.	n_z	0.09
1024	b-h	#_9	1
1025	m-h	#_^	0.52
1026	&-I	#_n	0.28
1026	I-0	?_.	0
1026	.-r	I_l	0.03
1026	.-0	p_.	0.04
1026	.-r	._#	0.03
1027	d-l	#_4	0.34
1028	J-Z	i_0	0.31
1029	O-u	l_r	0.29
1029	r-.	O_d	0.04
1030	t-.	1_s	0.07
1031	9-3	l_r	0.29
1031	r-.	9_n	0.04
1032	.-l	#_I	0.06
1033	I-0	m_l	0
1033	.-s	0_#	0.08
1034	.-v	8_.	0.34
1034	.-0	._r	0.04
1035	d-.	n_z	0.09
1036	m-n	#_4	0.14
1037	.-d	u_z	0.11
1038	J-.	#_5	0.52
1039	r-p	#_4	0.52
1040	?-k	r_I	0.66
1041	l-r	0_.	0.13
1041	.-z	l_#	0.25
1042	.-p	#_r	0.16
1043	k-p	#_^	0.24
1044	^-u	r_b	0.28
1045	U-^	r_m	0.28
1046	n-.	0_I	0.1
1046	I-.	n_N	0.11
1046	N-n	I_z	0.14
1047	p-v	r_0	0.66
1048	t-s	#_r	0.34
1048	r-n	t_4	0.29
1049	t-n	s_I	0.34
1050	r-.	O_k	0.04
1051	.-s	#_.	0.08

1051	.-t	._r	0.07
1052	4-i	t_l	0.27
1053	k-.	N_#	0.18
1054	f-t	#_u	1
1055	.-t	f_.	0.07
1055	.-s	._#	0.08
1056	&-^	#_.	0.28
1056	.-N	&_k	0.38
1057	t-.	#_9	0.07
1058	E-&	v_s	0.19
1059	l-v	#_I	0.44
1060	9-I	v_r	0.29
1060	r-.	9_J	0.04
1061	k-d	r_#	0.52
1062	b-w	#_9	0.66
1063	f-j	#_u	1
1064	0-1	?_#	0.29
1065	w-v	i_i	0.08
1066	h-.	l_i	0.07
1067	.-t	k_#	0.07
1068	&-4	#_n	0.27
1069	9-7	#_r	0.3
1070	.-d	n_z	0.11
1071	i-4	b_t	0.28
1072	t-d	i_#	0.09
1073	.-r	i_#	0.03
1074	9-1	g_r	0.29
1074	r-.	9_l	0.04
1075	.-b	#_E	0.21
1076	m-b	#_1	0.44
1077	t-s	k_#	0.34
1078	r-l	b_i	0.11
1079	5-2	l_z	0.29
1080	p-b	#_r	0.14
1081	p-b	#_U	0.14
1082	g-k	#_a	0.11
1082	a-O	g_?	0.25
1083	.-0	k_w	0.04
1083	w-v	._O	0.08
1084	S-C	#_&	0.19
1085	l-k	#_a	0.66
1086	p-k	#_r	0.23
1087	g-k	#_.	0.11
1087	.-r	g_5	0.03
1088	.-k	#_l	0.1
1088	l-r	._^	0.13
1089	O-^	k_l	0.28
1090	D-d	#_0	0.05
1090	0-I	D_p	0.01
1091	D-d	#_0	0.05
1091	0-I	D_p	0.01
1092	k-d	#_r	0.52
1093	l-.	#_9	0.08
1094	b-.	#_i	0.21
1095	.-i	#_l	0.13
1096	1-E	#_l	0.28
1097	.-i	#_l	0.13
1098	I-E	#_n	0.21
1098	D-.	5_z	0.13
1099	I-E	#_n	0.21
1099	i-2	d_d	0.3
1099	d-.	i_#	0.09
1100	h-.	#_&	0.07
1100	&-E	h_v	0.2
1100	0-I	v_d	0.01
1100	&-0	d_n	0.24
1101	.-d	l_#	0.11
1102	.-l	f_4	0.06
1103	9-3	f_r	0.29
1103	r-.	9_s	0.04
1104	0-O	f_r	0.27
1104	g-k	r_.	0.11
1104	.-l	g_I	0.06
1104	v-f	I_.	0.14
1104	.-t	v_#	0.07
1105	t-f	#_r	0.52
1106	l-r	g_5	0.13
1107	b-g	#_l	0.31
1107	l-r	b_5	0.13
1108	t-d	r_#	0.09
1109	.-h	#_j	0.08
1110	.-I	#_w	0.1
1110	w-m	._9	1
1111	h-.	#_I	0.07
1111	O-0	m_r	0.25
1111	r-.	O_l	0.04
1112	n-m	I_p	0.11
1113	E-4	h_l	0.26
1114	&-I	h_b	0.28
1115	m-n	I_&	0.14
1115	&-a	m_k	0.18
1115	I-4	l_t	0.28
1116	0-I	#_.	0.01
1116	.-n	0_s	0.1
1117	0-.	J_&	0.07
1117	v-b	&_#	0.18
1118	C-J	#_5	0.18
1119	1-&	l_k	0.22
1120	f-T	&_#	0.44
1121	v-s	i_#	0.66
1122	m-l	#_E	0.38
1123	t-l	#_a	0.38
1124	n-m	#_4	0.11
1125	r-n	#_i	0.29
1126	O-0	d_l	0.25
1127	t-d	a_#	0.09
1128	.-p	#_4	0.16
1129	b-p	#_a	0.16
1130	&-a	p_.	0.18
1130	.-r	&_s	0.03
1131	0-.	#_p	0.07
1131	a-O	p_n	0.25
1132	j-p	#_a	1
1133	l-p	#_^	1
1134	p-.	#_r	0.24
1135	.-r	#_4	0.03
1136	j-.	z_0	0.09
1137	w-r	#_i	0.21
1138	m-r	#_i	0.66
1139	k-p	&_s	0.24
1140	E-i	p_n	0.28
1140	n-.	E_t	0.1
1141	u-3	r_l	0.3
1142	5-u	r_m	0.27
1143	S-s	#_i	0.17
1144	T-s	#_I	0.25
1145	.-s	#_k	0.08
1146	t-s	#_.	0.34
1146	.-l	t_4	0.06
1147	.-l	s_I	0.06
1148	O-5	n_r	0.26
1148	r-.	O_#	0.04
1149	T-s	#_.	0.25
1149	.-t	T_I	0.07
1150	.-r	t_4	0.03

1151	.-s	#_.	0.08
1151	.-t	._l	0.07
1151	l-r	._1	0.13
1152	4-&	t_k	0.27
1153	s-T	#_I	0.11
1154	s-T	#_^	0.11
1155	w-.	t_I	0.13
1156	k-t	#_r	0.16
1157	^-I	r_s	0.27
1158	a-^	#_n	0.26
1158	D-d	n_0	0.05
1159	0-^	#_n	0.28
1159	D-d	n_.	0.05
1159	.-j	D_0	0.1
1160	w-v	#_E	0.08
1160	t-?	n_I	0.21
1160	I-0	t_l	0
1161	w-v	#_E	0.08
1161	?-C	n_0	0.66
1161	.-r	0_#	0.03
1162	t-d	n_#	0.09
1163	r-w	#_4	0.14
1164	r-w	#_E	0.14
1165	S-w	#_1	1
1166	.-t	l_#	0.07
1167	O-0	w_r	0.25
1168	u-.	j_r	1
1168	r-.	8_a	0.04
1168	a-O	r_n	0.25
1169	.-j	#_9	0.1
1170	r-j	#_a	0.66
1170	D-d	n_0	0.05
1171	s-z	#_I	0.07
1172	.-d	n_s	0.11
1173	s-S	#_E	0.09
1173	E-7	s_r	0.29
1174	r-.	f_i	0.04
1175	T-t	#_r	0.18
1175	5-^	r_s	0.29
1175	.-0	5_n	0.04
1176	?-C	a_0	0.66
1177	.-r	O_k	0.03
1178	.-l	b_5	0.06
1179	s-J	#_I	0.38
1180	i-I	#_.	0.15
1180	.-n	i_v	0.1
1181	h-.	#_5	0.07
1181	I-0	p_n	0
1182	D-z	#_0	0.34
1182	0-i	D_b	0.24
1182	a-0	r_#	0.26

Appendix E: Main Corpus of Target-Pun Pairs

The following table lists the 1182 target pun pairs selected from Sobkowiak's (1992) corpus, including their transcription (p.phon, t.phon(, idiomaticity, page number (p.) form Crosbie (1977), grammaticality (gram.) of pun (P)/target (T)/both(B), notes, language/dialect peculiarities, CLAWS tagging for pun (C P) and target (C T), a marking field that ended up not being used (x), sum cost (sum) according to the preliminary cost table, sum cost per segments (sum/phon), stress shift (e=earlier in pun than in target, l=later), and syllable number (-1=one syllable less in pun than in target, +1=one more, +2=two more).

no.	pun	targ.	p. phon	t. phon.	idiomaticity	p.	gram.	note	dialect	C P	C T	x	sum	sum/phon.	str.	syll #
1	hence	hens	hEns	hEnz		110	T	what?		CJS	NN2		0.07	0.0175		
2	neither	nether	niD0r	nED0r	coll. nether world	158	T			PNI	NN1	R	0.28	0.056		
3	and loaf	alone	0nl5f	0.l5n	idiom live by bread alone	140	B			CJS NN1	AV0	E+	0.76	0.152		
4	dun her	dinner	d^n0r	dIn0r	title man who came to dinner (Moss)	71	B			VVB PNP	NN1	Q+	0.29	0.058		
5	kama	camel	kam0.	k&m0l	coll. made of camel hair	128	T			UNC	NN1	I	0.26	0.052		
6	kith	kiss	kIT	kIs	idiom kissing cousins	130	T			NN1	NN1	C+	0.11	0.036666665		
7	miaow	me how	mi.2	mih2		149	T			ITJ	PNP AVQ		0.07	0.0175		
8	N	end	En.	End	coll. end man	156	T	see also Nero		ZZ0	NN1	R	0.09	0.03		
9	spoorious	spurious	sp.8ri0s	spjuri0s	coll. spurious claim	217	B			AJ0	AJ0	C	0.38	0.0475		
10	vasted	vested	v&stId	vEstId	coll. vested interest	243	T			AJ0	AJ0	C	0.19	0.031666666		
11	abdominal	abominable	0bdamIn..l	0b.amIn0bl	coll. abominable snowman	5	B			AJ0	AJ0	C	0.39	0.039		
12	acidic	I see Dick	0sIdIk	1sidIk		7	ne			AJ0	PNP VVB NP0		0.35	0.058333334	e	
13	acoustic	a cue stick	0k.ustIk	0kjustIk	coll. cue stick	7	ne			AJ0	AT0 NN1 NN1	R	0.09	0.01125		
14	adamant	Adam and	&d0m0nt	&d0m0nd	coll. Adam and Eve	7	ne			AJ0	NP0 CJC	B	0.11	0.015714286		
15	alike	I like	0l1k	1l1k	idiom I like to think so	10	T	adnominatio		AJ0	PNP VVB	I+	0.2	0.05		
16	alimentary	elementary	&l0mEn?0rI	El0mEn?0rI	quote elementary, my dear watson	10	B	what?		AJ0	AJ0	Q+	0.19	0.019		
17	artesian	a-teasin'	0rtiZ0n	0.tiz0n		16	T	a-Xing prefix		AJ0	VVG		0.41	0.058571428		
18	awkward	hawk ward	.Okw0rd	hOkwOrd	_def.	20	ne			AJ0	NN1 NN1		0.32	0.045714285		
19	backward	back ward	b&kw0rd	b&kwOrd	_def.	22	ne			AJ0	NN0 NN1		0.25	0.035714287		
20	bellicose	belly 'cause	bElIk5s	bElIkOz		28	T			AJ0	NN1 CJC		0.33	0.047142856	l	
21	bellicose	very close	bElIk.5s	vErIkl5s	_def.	28	ne		Chinese	AJ0	AV0 AJ0		0.37	0.04625		
22	benign	be nine	bIn1n	bin1n	_def.	28	ne			AJ0	VBI CRD		0.15	0.03		
23	boisterous	boys to us	b3st0r0s	b3zt0.0s		32	T			AJ0	NN2 PRP PNP		0.1	0.0125		
24	bonkers	yonkers	baNk0rz	jaNk0rz	_def.	32	ne			AJ0	NP0		0.66	0.09428571		
25	brief	beef	brif	b.if	coll. beefeater	35	B			AJ0	NN0	R	0.03	0.0075		
26	chary	cherry	C7rI	CErI	coll. cherry picker	46	B			AJ0	NN1	R	0.29	0.0725		
27	chic	cheek	Sik	Cik	idiom tongue in cheek	47	T			AJ0	NN1	I+	0.15	0.05		
28	coined	kind	k3nd	k1nd	idiom my kinda girl	51	T			AJ0	NN1	I+	0.28	0.07		
29	comical	chemical	kamIk0l	kEmIk0l	coll. chemical combination	52	B			AJ0	AJ0	C	0.27	0.03857143		
30	consecrated	concentrated	kans0.kr4tId	kans0ntr4tId	coll. concentrated lye	54	B			AJ0	AJ0	C	0.24	0.02		
31	contrite	can't write	k0ntr1t	k&ntr1t		54	T			AJ0	VM0 NN0 VVB		0.24	0.034285713		
32	covert	cover	k5v9rt	k^v0r.	idiom from cover to cover	57	T	eye-pun		AJ0	NN1	E+	0.64	0.10666667	l	
33	curly	pearly	k9rlI	p9rlI	coll. Pearly Gates	60	B			AJ0	AJ0	R	0.23	0.046		
34	curtain	certain	k9r?0n	s9r?0n	idiom a certain something	61	B			AJ0	AJ0	T+	0.27	0.045		
35	disgusting	this gusting	dIsg^stIN	DIsg^stIN		67	T	no 2nd script		AJ0	DT0 AJ0		0.05	0.0055555557		
36	dismal	decimal	dIz.m0l	dEsIm0l	_def.	68	B			AJ0	NN1		0.38	0.054285713		
37	dour	our	d20r	.20r	title Oh God, our help...	69	B			AJ0	DPS	Q+	0.11	0.0275		
38	elementary	alimentary	El0mEnt0rI	&l0mEnt0rI	coll. alimentary canal	75	B			AJ0	AJ0	C	0.2	0.02		
39	erotic	erratic	0ra?Ik	0r&?Ik	coll. erratic behavior	78	B			AJ0	AJ0	C	0.18	0.03		
40	erroneous	erogenous	0r5..ni0s	0raJ0n.0s	coll. erotic zones	78	B	little diff!!??		AJ0	AJ0	C	0.99	0.11		
41	eternal	internal	i.t9rn0l	Int9rn0l	coll. internal medicine	79	B			AJ0	AJ0	C	0.2	0.025		
42	evanescent	if it isn't	Ev0nEs0nt	If0dIz0nt	idiom if it isn't my old pal	80	T			AJ0	CJS PNP VBZ XX0	I+	1.04	0.115555555		
43	evil	he will	.ivIl	hiwIl		80	T			AJ0	PNP VM0		0.15	0.03		

44	fallow	fellow	f&l5	fEl5	coll. fellow Americans	84	B			AJ0	AJ0	C	0.19	0.0475		
45	fecund	second	fik0nd	sEk0nd	coll. second nature	85	B			AJ0	ORD	C	0.55	0.09166667		
46	flair	fair	fl7r	f.7r	title my fair lady	88	B			AJ0	AJ0	Q+	0.06	0.015		
47	fortunate	fortune	fOrC0n0t	fOrC0n..	coll. fortune cookie	92	B			AJ0	NN1	R	0.11	0.01375		+1
48	fragrant	flagrant	fr4gr0nt	fl4gr0nt	coll. flagrant error	92	B			AJ0	AJ0	C	0.13	0.01625		
49	froggy	foggy	fragI	f.agI	title A Foggy Day in L. Town (Gershwin)	93	B			AJ0	AJ0	Q+	0.03	0.006		
50	frozen	chosen	fr5z0n	C.5z0n	idiom God's chosen people	94	B			AJ0	AJ0	Q+	1.03	0.17166667		
51	fundamental	from the mantel	f.^nd0mEn?0l	fr0mD0m&n?0l		94	T	what?		AJ0	PRP AT0 NN1		0.71	0.059166666		
52	hard	heart	hard	hart	proverb home is where the heart is	107	T	what?		AJ0	NN1	S+	0.09	0.0225		
53	hollow	hello	hal5	hEl5		112	ne			AJ0	ITJ		0.27	0.0675		
54	impotent	important	Imp0.t0nt	ImpOr?0nt	idiom important moment in history	118	B			AJ0	AJ0	I+	0.46	0.05111111	e	
55	insufferable	inseparable	Ins^f0r0b0l	InsEp0r0b0l	coll. inseperable friends	120	B			AJ0	AJ0	C	0.49	0.044545453		
56	invisible	visible	InvIs0b0l	..vIs0b0l	idiom no visible means of support	121	B	# of syll. diff.		AJ0	AJ0	I+	0.2	0.022222223		+1
57	juvenile	javelin	Juv0n1l	J&v.lIn	_def.	127	T	# of syll. diff.		AJ0	NN1		1.16	0.16571428		
58	lesbian	let's be an	lE.zbi0n	lEtsbi0n		136	T			AJ0	VVB PNP VBI AT0		0.14	0.0175		
59	marrowing	harrowing	m&r5wIN	h&r5wIN	coll. a harrowing experience	147	B			AJ0	AJ0	C	0.38	0.054285713		
60	meager	eager	mig0r	.ig0r	proverb an early winter makes ... ?	148	B			AJ0	AJ0	S+	0.14	0.028		
61	menstrual	minstrel	mEnstr0l	mInstr0l	title lay of the last minstrel (Walter Scott)	148	T			AJ0	NN1	Q+	0.21	0.02625		+1
62	morose	rose	m0r5s	..r5z	coll. rose-colored glasses	153	T	# of syllables different		AJ0	NN1	C	0.25	0.05		+1
63	nasal	naval	n4z0l	n4v0l		157	B			AJ0	AJ0		0.52	0.104		
64	natty	naughty	n&?I	nO?I	coll. naughty boy	157	B			AJ0	AJ0	C	0.27	0.0675		
65	nautical	naughty	nO?Ik0l	nO?I...	idiom that's very naughty of them	157	B	# of syllables different		AJ0	AJ0	I+	0.2	0.028571429		+1
66	nee	non	ni.	nan	coll. non plus ultra	158	T	morpheme only		AJ0	UNC	F	0.39	0.13		
67	nixed	mixed	nIkst	mIkst	coll. mixed emotions	160	B			AJ0	AJ0	C	0.14	0.028		
68	numb	dumb	n^m	d^m	coll. dumb blonde	162	B			AJ0	AJ0	C	0.38	0.12666667		
69	odd	old	a.d	5ld	proverb old soldiers never die	164	B			AJ0	AJ0	S+	0.33	0.11		
70	opportune	opera tune	ap0r.tjun	ap.r0tjun	coll. opera tune	165	ne			AJ0	NN0 NN1	R	2	0.25	l	
71	orange	arrange	OrInJ	0r4nJ	coll. arrange one's life	119	T	see indigo		AJ0	VVB	I	0.55	0.11	e	
72	passable	possible	p&s0b0l	pas0b0l	proverb politics is the art of the possible	171	B			AJ0	AJ0	S+	0.23	0.032857142		
73	penal	penile	pin0l	pin1l	_def.	173	B	WS had penial-penal		AJ0	AJ0		0.2	0.04		
74	platonic	tonic	pl0tanIk	...tanIk		178	T	odd, just □ inserted		AJ0	NN1		0.26	0.0325		+1
75	polite	lighted	p0l1t..	..l1?Id	to make a polite entrance	178	B	entrance: door vs. entering		AJ0	AJ0		0.57	0.08142857	l	
76	prostituting	prosecuting	prast0tjutIN	pras.0kjutIN	coll. prosecuting attorney	183	B			AJ0	AJ0	R	0.23	0.019166667		
77	pubic	public	pjub.Ik	p.^blIk	coll. public-opinion poll	184	B			AJ0	AJ0	R	0.46	0.065714285		
78	pubic	cubic	pjubIk	kjubIk	coll. cubic centimeter	184	B	sentimeter?		AJ0	AJ0	C	0.24	0.04		
79	quaint	ain't	kw4nt	..4nt	idiom appears to be what it ain't	188	T			AJ0	VBB XX0	E+	0.21	0.042		
80	quibbling	sibling	kwIblIN	s.IblIN	coll. sibling rivalry	188	B			AJ0	NN1	R	0.38	0.054285713		
81	quipping	whipping	kwIpIN	.wIpIN	coll. whipping boy	189	B			AJ0	AJ0	R	0.1	0.016666668		
82	rare	pear	r7r	p7r	quote a partrigde in a pear tree	218	B	see spree		AJ0	NN1	Q+	0.66	0.22		
83	reclining	declining	rIkl1nIN	dIkl1nIN	coll. the declining years	193	B			AJ0	AJ0	C	0.44	0.055		
84	rhymed	rind	r1md	r1nd	coll. orange rind	196	B			AJ0	NN1	R	0.11	0.0275		
85	rid	red	rId	rEd	coll. red-letter day	197	B			AJ0	AJ0	R	0.2	0.06666667		
86	rile	real	r1l	ril		197	B		Irish	AJ0	AJ0		0.3	0.1		
87	rued	road	rud	r5d		199	B			AJ0	NN1		0.27	0.09		
88	runed	ruined	ru.nd	ruInd		200	B			AJ0	AJ0		0.11	0.022		
89	short	sort	SOrt	sOrt	coll. that sort of thing	210	T	*Ident(man) bad		AJ0	NN1	I	0.09	0.0225		
90	sinister	spinster	s.InIst0r	spIn.st0r	_def.	211	ne			AJ0	NN1		0.34	0.037777778		+1
91	snaky	shaky	sn4kI	S.4kI		214	B			AJ0	AJ0		0.27	0.054		
92	sordid	assorted	.sOrdId	0sOrtId	coll. assorted flavors	215	B			AJ0	AJ0	C	0.16	0.022857143		-1
93	sour	our	s20r	.20r	idiom ... is our heritage	216	B			AJ0	DPS	E+	0.08	0.02		
94	stationary	station	st4S.n7rI	st4S0n...	coll. station wagon	219	B			AJ0	NN1	R	1.2	0.13333334		+2
95	terse	Thurs(day)	t9rs	T9rz		229	B		dialect?	AJ0	NP0		0.25	0.0625		
96	tintative	tentative	tIn?0tIv	tEn?0tIv		119	B	P non-word		AJ0	AJ0		0.2	0.025		
97	tortuous	thought she was	tOrCu.0s	TOtSiw0z		234	T	Ident(voic)-Sc/V_#		AJ0	VVD PNP VBD		1.21	0.15125		
98	touring	towering	t8rIN	t2rIN	idiom mountains towering among the clouds	234	B			AJ0	AJ0	E+	0.3	0.06		
99	tropical	topical	trapIk0l	t.apIk0l	coll. topical subject	236	B	what?		AJ0	AJ0	C	0.03	0.00375		
100	utter	butter	.^?0r	b^?0r	coll.s bread and butter utter nonsense	241	B	collocation interaction		AJ0	NP0	B	0.21	0.042		
101	vapid	rapid	v&pId	r&pId	coll. rapid transit system	243	B			AJ0	AJ0	R	0.44	0.088		
102	vast	west	v&st	wEst	coll. West Point	243	B			AJ0	NN1	R	0.27	0.0675		
103	half vast	half-assed	h&f&st	h&f&sd	coll. half-assed	243	B			AJ0	AJ0	R	0.11	0.018333333		
104	vicarious	why carry us	v1k7ri0s	w1k7rI0s		245	T			AJ0	AVQ VVB PNP		0.18	0.0225		
105	vicious	wishes	vIS0s	wIS0z	idiom best wishes for a merry christmas	245	T			AJ0	NN2	I+	0.15	0.03		

106	vocal	local	v5k0l	l5k0l	coll. local anesthetic
107	vulgar	Volga	v^lg0r	valg0.	title the Volga boatman
108	vying	dying	v1IN	d1IN	coll. a dying breed
109	weird	word	w6rd	w9rd	idiom word of mouth
110	wrangled	spangled	r.&Ng0ld	sp&Ng0ld	coll. star-spangled banner
111	a cute	acute	0kjut	0kjut	
112	not T.	naughty	nati	nO?I	
113	old boys	oboes	5ldb3z	5..b5z	
114	Indian pendants	independence	Indi0npEnd0ns	Ind.0.pEnd0ns	coll. declaration of independence
115	one's cone	one stone	w^nsk5n	w^nst5n	proverb kill two birds with one stone
116	past man	postman	p&stm&n	p5stm0n	title the postman always rings twice
117	carafe tea	crafty	k0r&fti	k.r&ftI	
118	sore shearer	sorcerer	sOrS6r0r	sOrs0r0r	coll. sorcerer's apprentice
119	amphibian	I'm fibbing	0mfIbi0n	1mfIb.0N	idiom you think I'm fibbing, do you?
120	farther	father	farD0r	fa.D0r	coll. step father
121	fit	fifth	fI.t	flfT	
122	innuendo	in the window	InjuwEnd5	InD0wInd5	
123	internal	eternal	0nt9rn0l	0.t9rn0l	title hope springs eternal
124	lying	lion	l1IN	l10n	coll. Richard the Lionhearted
125	right	night	r1t	n1t	title I could have danced all night
126	sad	sat	s&d	s&t	
127	violetly	violently	v10lI.tlI	v10l0ntlI	
128	wound	round	w2nd	r2nd	idiom see you around
129	afire	of fire	0f1r	Of1r	coll. pillar of fire
130	mine eyes	mayonnaise	m1.n1z	m40n4z	_def.
131	blight	bite	bl1t	b.1t	idiom his bark is worse than his bite
132	dink	link	dINk	lINk	coll. weakest link
133	estrangers	strangers	Estr4nJ0rz	.str4nJ0rz	title strangers in the night
134	par	pa	par	pa.	
135	each	itch	iC	IC	_def.
136	a llama	alarm her	0la.m0.	0larm0r	
137	aloof	a roof	0luf	0ruf	_def.
138	a nymph	(a)n imp	0nImf	0nImp	hare lip cleft palate talk
139	a tea daze	80 days	0tid4z	4?Id4z	title around the world in 80 days
140	a ghost	aghast	0g5st	0gast	idiom be aghast at something
141	a swan	as one	0swan	0sw^n	title two hearts that beat as one
142	a tome	at home	0t.5m	0th5m	idiom feel at home
143	a goon	a-gone	0gun	0gan	title Gone with the Wind
144	a Freud	afraid	0fr3d	0fr4d	
145	a boot	about	0but	0b2t	idiom it's about time
146	an oaf	a loaf	0n5f	0l5f	proverb half a loaf is better than none
147	avaunt	I want	0vOnt	1want	
148	Begorra	Gomorrah	b0gOr0	g0mOr0	coll. Sodom and Gomorrah
149	wow	how	w2	h2	quote how do I love thee
150	pickle	tickle	pIk0l	tIk0l	idiom tickle one's fancy
151	ream	dream	.rim	drim	title to sleep, perchance to dream
152	yank	thank	j&Nk	T&Nk	idiom have only oneself to thank
153	yell	hell	jEl	hEl	idiom go to hell
154	zinc	think	zINk	TINk	idiom stop to think
155	aberration	adoration	&b0r4S0n	&d0r4S0n	coll. mutual adoration
156	acre	a cur	4k0r	0k9r	_def.
157	acrimony	matrimony	.&krIm5nI	m&trIm5nI	_def.
158	acrimony	alimony	&krIm5nI	&.lIm5nI	_def.
159	adder	had 'er	.&d0r	h&d0r	idiom I had her
160	ale	hell	.4l	hEl	idiom what the hell
161	alternative	altar native	Olt9rn0tIv	Olt0rn4tIv	
162	amnesty	I'm nasty	&mn0stI	1mn&stI	
163	antipasto	anti-pasta	&n?Ipast5	&n?Ipast0	_def.
164	apostle	a parcel	0pa.s0l	0pars0l	_def.
165	argon	Oregon	ar.g0n	Or0g0n	_def.
166	arson	horsin'	.ars0n	hOrs0n	idiom stop horsing around
167	aspen	askin'	&sp0n	&skIn	

246	B			AJ0	AJ0	C	0.44	0.088	
247	B			AJ0	NP0	Q+	0.29	0.048333332	
247	B			AJ0	AJ0	C	0.44	0.11	
251	T			AJ0	NN1	I+	0.3	0.075	
256	B			AJ0	AJ0	C	0.9	0.1125	
186	B	see pyorrhea		AT0 AJ0	AJ0		0	0	
225	B			XX0 ZZ0	AJ0		0.51	0.1275	
245	B	see vile		AJ0 NN2	NN2		0.46	0.07666667	
172	B	see pendant		AJ0 NN1	NN1	I	0.23	0.017692307	
53	B			PNP POS NN1	CRD NN1	S+	0.14	0.02	
171	B			AJ0 NN1	NN1	Q+	0.52	0.074285716	
43	B			AJ0 NN0	AJ0		0.14	0.02	
215	B			AJ0 NN1	NN1	R	0.39	0.04875	l
12	B	Adv -> subordinate phrase		AV0	PNP VBB VVG	E+	0.47	0.05875	
84	P	Target not synt□,3 ???		AV0	NN1	R	0.03	0.005	
88	B			AV0	AV0		0.62	0.155	
120	T			AV0	PRP AT0 NN1		1.47	0.16333333	
121	ne			AV0	AV0	Q+	0.1	0.0125	
256	B	see wretched		AV0	NN1	I	0.2	0.05	
197	B			AV0	NN1	Q+	0.31	0.10333333	
201	B			AV0	VVD		0.09	0.03	
118	B	see indigo		AV0	AV0		0.11	0.012222222	
255	B			AV0	AV0	I+	0.14	0.035	
176	B	see pillow		AV0	PRP NN1	I	0	0	
150	ne			AJ0 NN2	NN1		0.61	0.10166667	
31	B			NN1	NN1	E+	0.06	0.015	
67	B			NN1	NN1	C	0.44	0.11	
79	B			NN2	NN2	Q+	0.38	0.038	+1
169	B	WS had par-pear		UNC	NN1		0.03	0.01	
73	ne	no 2nd script		AT0	NN1		0.1	0.05	
140	T			AT0 NN1	VVB PNP		0.08	0.011428571	
11	ne		Chinese	UNC	AT0 NN1		0.11	0.0275	
162	B			AT0 NN1	AT0 NN1		0.23	0.046	
63	B			AT0 NN0 NN1	CRD NN2	Q+	0.55	0.09166667	l
97	B			AT0 NN1	AJ0	I+	0.27	0.054	
223	B	word boundaries differ p->t		AT0 NN1	PRP PN1	Q+	0.21	0.042	
233	B	word boundaries diff. p->t		AT0 NN1	PRP NN1	I+	0.07	0.014	
100	B	title mangled, ouch		AT0 NN1	VVN	Q+	0.29	0.0725	
93	T			AT0 NP0	AJ0		0.29	0.058	
33	T			AT0 NN1	AV0	E+	0.29	0.0725	
163	B			AT0 NN1	AT0 NN1	S+	0.23	0.0575	
19	T		French	ITJ	PNP VVB		0.53	0.106	
27	T	#NAME?		ITJ	NP0	B	1.1	0.18333334	
256	ne			ITJ	AVQ	Q+	0.38	0.19	
175	B			VVI	VVI	I+	0.18	0.036	
192	B			VVI	VVI	Q+	0.09	0.0225	
259	B			VVI	VVI	I+	1	0.25	
259	B			VVI	NN1	E+	0.44	0.14666666	
262	B			VVI	VVI	I+	0.66	0.165	
6	B			NN1	NN1	C	0.21	0.02625	
7	ne			NN1	AT0 NN1		0.57	0.1425	e
7	ne			NN1	NN1		0.33	0.036666665	
7	ne			NN1	NN1		0.23	0.02875	
7	T	see boa		NN1	VHD PNP	I+	0.07	0.014	
10	T	see ale-all, see rum		NN1	NN1	I+	0.33	0.11	
11	ne			NN1	NN0 NN1		0.57	0.057	l
12	ne			NN1	PNP VBB AJ0		0.46	0.065714285	e
14	ne		Italian	NN1	NN1		0.23	0.025555555	
14	ne			NN1	AT0 NN1		0.04	0.0057142857	
15	ne			NN1	NP0		0.31	0.051666666	
16	T			NN1	VVG	E+	0.31	0.051666666	
17	T	see poplar, alder		NN1	VVG		0.24	0.048	

168	ass	grass	..&s	gr&s	title the green, green grass of home	17	B			NN1	NN1	Q+	0.42	0.105		
169	ax	X	&ks	Eks	_def.	20	ne			NN1	ZZ0		0.19	0.06333333		
170	badger	bad you	b&J0r	b&J0.		22	T			NN1	AV0 PNP		0.03	0.006		
171	badinage	bad in age	b&dInaZ	b&dIn4J		22	T			NN1	AV0 PRP NN1		0.6	0.08571429		
172	bagel	bay gull	b4g0l	b4g^l	_def.	22	ne			NN1	NN0 NN1		0.27	0.054		
173	banana	been any	b0n&n0	bInEnI	idiom been any messages for me	23	T			NN1	VBN AT0	I+	0.19	0.031666666	l	
174	bard	bird	bard	b9rd	proverb birds of a feather ...	24	B			NN1	NN1	S+	0.29	0.0725		
175	bargain	bar gain	barg0n	barg4n	_def.	24	ne			NN1	NN0 NN1		0.28	0.046666667		
176	beck	back	bEk	b&k		26	B	what?		NN1	NN1		0.2	0.06666667		
177	beguine	begin	bIgin	bIgIn	idiom begin at the beginning	28	T	beguine=dance		NN1	VVI	S+	0.1	0.02		
178	beret	bray	b0r4	b.r4	_def.	28	ne			NN1	NN1		0.04	0.01		+1
179	berry	bury	bErI	b9rI	title bury me on the lone prairie	28	T			NN1	VVB	Q+	0.29	0.0725		
180	billet	Billy	bIIIt	bIII.	coll. hill billy	30	ne	proper name		NN1	NP0	R	0.07	0.014		
181	bloomer	blew my	blum0r	blum1.		31	T			NN1	VVD DPS		0.23	0.038333334		
182	bone	born	b5.n	bOrn		32	ne		s. US	NN1	VVB		0.3	0.075		
183	bordello	bored hello	bOrd.El5	bOrdhEl5	_def.	33	ne			NN1	AJ0 NN1		0.07	0.00875	l	
184	burro	borrow	b9r5	bar5		38	T			NN1	VVI		0.29	0.0725		
185	Carnage	Carthage	karnIJ	karTIJ		0	ne	where?		NN1	NP0		1	0.16666667		
186	carrot	car ought	k&r0t	karOt	proverb put the cart before the horse	43	T			NN1	NN1 VMB	S+	0.48	0.096		
187	caucus	carcass	kO.k0s	kark0s	_def.	44	ne			NN1	NN1		0.29	0.048333332		
188	centaur	cent whore	sEnt.Or	sEnthOr	_def.	45	ne			NN1	NN0 NN1		0.07	0.01		
189	chair	cheer	C7r	Cir	idiom three hearty cheers	45	B			NN1	NN1	I+	0.29	0.096666664		
190	chairman	charming	C7rm0n	CarmIN	idiom it's been charming	45	T			NN1	AJ0	I+	0.43	0.071666665		
191	chandelier	Chanticleer	S&nd0.l6r	C&n?0kl6r	_def.	46	ne			NN1	NP0		0.45	0.05		
192	chivalry	shivery	SIv0lrI	SIv0.rI	_def.	48	ne			NN1	AJ0		0.06	0.008571428		
193	chops	cops	Caps	kaps		48	T			NN2	NN2		0.38	0.095		
194	chowder	showed her	C2d0r	S5d0r		48	T			NN1	VVD PNP		0.48	0.096		
195	chub	shove	C^b	S^v	idiom show someone around	48	T			NN1	VVI	I+	0.37	0.123333335		
196	commonplace	come and place	kam0npl4s	k^m0npl4s		52	T			NN1	VVB CJC VVB		0.21	0.023333333		
197	conversation	condensation	kanv0rs4S0n	kand0ns4S0n	_def.	54	ne			NN1	NN1		0.75	0.06818182		
198	cow	how	k2	h2	quote how, shall I love thee	57	ne			NN1	AVQ	Q+	0.38	0.19		
199	crash	cash	kr&S	k.&S		57	B	what?		NN1	NN1		0.03	0.0075		
200	crass	class	kr&s	kl&s	coll. middle-class	57	B	only morpheme		NN1	AJ0	R	0.13	0.0325		
201	creep	keep	krip	k.ip	coll. upkeep	58	B	only morpheme		NN1	NN1	R	0.03	0.0075		
202	criminal	cream in all	krIm0n0l	krimInOl		58	T			NN1	NN1 PRP AT0		0.4	0.05		
203	cudgel	cordial	k^.J0l	kOrJ0l		60	T			NN1	AJ0		0.32	0.053333335		
204	cymbal	simple	sImb0l	sImp0l	idiom not as simple as it looks	61	T			NN1	AJ0	E+	0.14	0.023333333		
205	Dallas	dollars	d&l0.s	dal0rz		229	ne	see Texas		NP0	NN2		0.34	0.05666667		
206	dance	dunce	d&ns	d^ns	_def.	63	ne	what?		NN1	NN1		0.27	0.0675		
207	deacon	beacon	dik0n	bik0n	_def.	63	ne			NN1	NN1		0.38	0.076		
208	deadlock	wedlock	dEdlak	wEdlak	title holy wedlock	63	ne	WS has deadlock-bedlock		NN1	NN1	Q+	1	0.16666667		
209	deceiving	receiving	dIsivIN	rIsivIN		63	ne			NN1	NN1		0.29	0.04142857		
210	dishes	this is	dISIz	DIsIz		68	B	knock-knock		NN2	DT0 VBB		0.14	0.028		
211	divan	divine	dIv&n	dIv1n	idiom look divine	68	T			NN1	AJ0	I+	0.22	0.044	e	
212	doormat	dormant	dOrm&.t	dOrm0nt	coll. lie dormant	69	T			NN1	AJ0	I	0.33	0.047142856		
213	druid	do it	druId	d.uIt		70	T			NN1	VDB PNP		0.12	0.024		
214	duenna thing	do anything	duEn0TIN	duEnITIN		71	T			NN1	VDI PNP		0	0		
215	dusk	dust	d^sk	d^st	_def.	72	ne			NN1	NN1		0.14	0.035		
216	dybbuk	the book	dIb0k	D0bUk	coll. book-of-the-month club	72	ne			NN1	AT0 NN1	C	0.34	0.068	e	
217	eagle	equal	ig.0l	ikw0l	idiom all men are created equal	73	T			NN1	AJ0	Q+	0.23	0.046		
218	ear	hear	.6r	h6r	coll. corn ear	73	T			NN1	VVB	R	0.07	0.023333333		
219	easel	weasel	.iz0l	wiz0l	coll. weasel out of a problem	74	T			NN1	VVB	I	0.13	0.026		
220	Easter	used to	.ist0r	just0.		74	T	WS had Easter-ceased to		NN1	VVD TO0		0.42	0.07		
221	echelon	it's a long	E.S0lan	Its0laN	title it's a long way to Tipperary	74	ne			NN1	PNP VBZ AT0 AJ0	Q+	0.51	0.07285714		
222	elastic	Alaska	0l&stIk	0l&sk0.	coll. Baked Alaska	75	B			NN1	NP0	C	0.27	0.03857143		
223	eon	a yawn	i.0n	0jOn	_def.	78	ne			NN1	AT0 NN1		0.58	0.145	e	
224	epidemic	epidermic	Ep0dE.mIk	Ep0d9rmIk	_def.	78	ne			NN1	NN1		0.33	0.036666665		
225	epitaph	epigram	Ep0t.&f	Ep0gr&m	_def.	78	ne			NN1	NN1		1.14	0.16285715		
226	erg	urge	9rg	9rJ	coll. urge s.o. on (phrasal verb)	78	T			NN1	VVB	P	0.66	0.22		
227	eucalyptus	you clipped us	juk0lIpt0s	juk.lIpt0s	idiom you clipped us	80	ne			NN1	PNP VVD PNP	I+	0.04	0.004		+1
228	F	if	Ef	If		83	T			NN1	CJS		0.21	0.105		
229	faker	acre	f4k0r	.4k0r	coll. God's acre	83	B			NN1	NN1	R	0.27	0.054		

230	famine	farming	f&.mIn	farmIN		84	B		N. Engl.	NN1	NN1		0.41	0.068333335		
231	feline	feelin'	fil1n	filIn	idiom feeling one's oats	85	T			NN1	VVG	E+	0.28	0.056		
232	finch	finished	fIn.C	fInIS	_def.	87	ne			NN1	AJ0		0.3	0.06		-1
233	fink	think	fINk	TINk	quote I think, therefore I am	87	T			NN1	VB	Q+	0.24	0.06		
234	fjord	Ford	fjOrd	f.Ord	_def.	88	ne			NN1	NP0		0.1	0.02		
235	flesh	fresh	flES	frES	quote the pause that refreshes	89	B	just morpheme		NN1	NN1	Q+	0.11	0.0275		
236	flotation	flirtation	fl5.t4S0n	fl9rt4S0n	_def.	89	ne			NN1	NN1		0.34	0.037777778		
237	folder	for older	f..5ld0r	f0r5ld0r	idiom show respect for older people	91	T	no 2nd script		NN1	PRP AJC	E+	0.11	0.01375		
238	font	found	fant	f2nd	_def.	91	T	see agate		NN1	AJ0		0.38	0.095		
239	frenzy	friends he	frEn.zI	frEndzi		93	T			NN1	NN1 PNP		0.24	0.034285713		
240	gaucho	Graucho	g.2C5	gr2C5	name Gaucho Marx	96	B			NN1	NP0	R	0.04	0.008		
241	gentile	gentle	JEnt1l	JEnt0l	quote not go gentle into ... (D. Thomas)	96	B			NN1	NN1	Q+	0.29	0.048333332		
242	gob	got	gab	gat	idiom if you gotta go ...	99	T			NN1	VMB	S+	0.44	0.14666666		
243	guitar	got to	gItar	ga?0.	idiom if you gotta go ...	103	T			NN1	VVN	S+	0.72	0.144		
244	gulch	gulls	g^lC	g^lz	_def.	103	ne	no 2nd script		NN1	NN2		1	0.25		
245	gulp	gulf	g^lp	g^lf	coll. gulf of Mexico	103	B			NN1	NN1	I	0.21	0.0525		
246	hail	hell	h4l	hEl		105	ne		Southern	NN1	NN1		0.26	0.086666666		
247	halo	hello	h4l5	hEl5	idiom hello there	106	T	greeting hello		NN1	ITJ	I+	0.26	0.065		
248	ham	am	h&m	.&m	idiom I am what I am	106	T			NN1	VBB	Q+	0.08	0.026666667		
249	hamlet	omelette	h&ml0t	.aml0t	_def.	106	ne			NN1	NN1		0.31	0.051666666		
250	harmony	hominy	harm0nI	ha.m0nI	_def.	107	ne			NN1	NN1		0.03	0.004285714		
251	hartebeest	heartbeat	har?0bist	hart.bi.t	idiom only a heartbeat away	108	B			NN1	NN1	E+	0.33	0.036666665		
252	helm	hem	hElm	hE.m	coll. hem and haw	110	T			NN1	VVI	B	0.06	0.015		
253	hornet	cornet	hOrn0t	kOrn0t		245	B	see vile		NN1	NN1		0.34	0.05666667		
254	horse	house	hOrs	h2.s	proverb a house divided against itself	113	ne			NN1	NN1	E+	0.33	0.0825		
255	hyena	high in a	h1in0	h1In0		115	T			NN1	AVO PRP AT0		0.1	0.02	l	
256	icon	acorn	1ka.n	4kOrn	_def.	117	ne			NN1	NN1		0.55	0.11		
257	icon	hike on	.1kan	h1kan		117	T			NN1	VVB AVP		0.07	0.014		
258	iguana	I ('m) gonna	Igwan0	1mgan0		117	T	distant, but syntax recov.		NN1	PNP VBB VMI TO0		1.94	0.32333332		
259	incest	insist	InsEst	InsIst	idiom well, if you insist	118	T			NN1	VVB	I+	0.21	0.035	e	
260	ingot	in God	Ingat	Ingad	idiom in god we trust	119	T			NN1	PRP NN1	Q+	0.11	0.022	e	
261	jacket	Jack ate	J&k0t	J&kEt		123	T			NN1	NP0 VVD		0.23	0.046	e	
262	jeep	cheep	Jip	Cip		124	T			NN1	AJ0		0.18	0.06		
263	jest	just	JEst	J^st		124	T			NN1	AV0		0.29	0.0725		
264	jigsaw	rickshaw	JIgsO	rIkSO	_def.	125	ne			NN1	NN1		0.65	0.13		
265	jog	jug	Jag	J^g	_def.	125	B	2nd script?		NN1	NN1		0.21	0.07		
266	juice	Jews	Jus	Juz	_def.	126	ne			NN1	NP0		0.07	0.023333333		
267	kalpa	calliper	k&l.p0.	k&lIp0r		128	ne	what?		NN1	NN1		0.15	0.021428572		-1
268	kangaroo	can grew	k&Ng0ru	k&Ng.ru		128	T			NN1	NN1 VVD		0.04	0.0057142857		+1
269	kibbitz	kibbutz	kIbIts	kIbuts		129	ne		Heb./Yid.	NN1	NN1		0.29	0.048333332		
270	kindle	candle	kInd0l	k&nd0l	_def.	129	ne		Yiddish	NN1	NN1		0.28	0.046666667		
271	kirsch	curse	k9rS	k9rs	_def.	130	ne			NN1	NN1		0.09	0.0225		
272	kitchen	catching	kIC0n	k&C0n	phrasal verb catch up	130	T			NN1	VVG	P+	0.28	0.056		
273	Latin	lettin'	l&?In	lE?In		134	T			NN1	VVG		0.19	0.038		
274	lava	lover	lav0.	lav0r	title lover come back to me	134	ne			NN1	PNP VHB PNP	Q+	0.04	0.008		
275	ledger	let your	lEJ0r	lEC0r	idiom let yourself go	136	T			NN1	VVB DPS	I+	0.18	0.036		
276	leprechaun	leprosy	lEpr0kOn	lEpr0sI.	_def.	136	ne			NN1	NN1		0.67	0.08375		
277	lessee	let's see	lE.si	lEtsi		136	T			NN1	VMB PNP VVI		0.07	0.014		
278	ligature	look what you're	lIg.0C0r	lUkw0C0r		138	T			NN1	VVB AVQ PNP VBB		0.52	0.065		
279	litmus	lit mouse	lItm0s	lItm2s	_def.	140	ne			NN1	AJ0 NN1		0.29	0.048333332		
280	lizard	wizard	lIz0rd	wIz0rd	title the electrical wizard	140	B			NN1	NN1	Q+	0.44	0.07333333		
281	locust	low curse	l5k0.st	l5k9rs.	_def.	141	ne			NN1	AJ0 NN1		0.4	0.057142857	e	
282	lorry	sorry	larI	sarI	title who's sorry now	142	T			NN1	AJ0	Q+	0.66	0.165		
283	magnet	maggot	m&gn0t	m&g.0t	_def.	145	B	no 2nd script		NN1	NN1		0.1	0.016666668		
284	malice	Alice	m&lIs	.&lIs	title Alice in Wondeland	145	B			NN1	NP0	Q+	0.14	0.028		
285	mandolin	Mandarin	m&nd0lIn	m&nd0rIn		146	ne		Chinese?	NN1	NN1		0.11	0.01375		
286	marijuana	marry one	m&rIwan0	m&rIw^n.	_def.	146	T			NN1	VVB PN1		0.25	0.03125	l	+1
287	marmot	mum ought	marm0t	m^.mOt	_def.	147	T			NN1	NN1 VMD		0.49	0.08166666		
288	mastodon	mustard on	m&st0.dan	m^st0rdan		147	T			NN1	NN1 PRP		0.31	0.034444444		
289	May	day	m4	d4	idiom rare as a day in June	147	B			NN1	NN1	E+	0.66	0.33		
290	mizzen	missin'	mIz0n	mIs0n		151	T			NN1	VVG		0.07	0.014		
291	momentum	memento	m5mEn?0m	m0mEn?5.	_def.	151	ne			NN1	NN1		0.62	0.0775		

292	monastery	monstery	man0st7rI	man.st0rI	_def.	152	ne			NN1	NN1		0.33	0.036666665		+1
293	mongoose	man-goose	maNgus	m&ngus		152	B			NN1	NN1		0.37	0.061666667		
294	moose	must	mus	m^s		152	T			NN1	VMB		0.28	0.093333334		
295	mortarfied	mortified	mOrtarf1d	mOr?I.f1d		153	T			NN1	AJ0		0.48	0.053333335		
296	moth	must	maT	m^s		153	T	please!		NN1	VM0		0.32	0.10666667		
297	mulch	much	m^lC	m^.C	idiom too much of a ...	154	T			NN1	AV0	E+	0.06	0.015		
298	muzzle	ma will	m^z0l	m^wI1		155	T			NN1	NN1 VMB		1	0.2	e	
299	newt	knew it	nu.t	nuIt	idiom noone knew	159	T			NN1	VVD PNP	I+	0.11	0.0275		-1
300	ninny	nanny	nInI	n&nI	_def.	159	ne			NN1	NN1		0.28	0.07		
301	ocelot	owes a lot	as0l0t	5z0lat		164	T			NN1	VVZ AT0 AJ0		0.58	0.096666664		
302	oil	I'll	3l	1l		165	T			NN1	PNP AM0		0.28	0.14		
303	omelet	amulet	am..l0t	&mjul0t	_def.	165	ne			NN1	NN1		1.27	0.18142857		
304	onomatopoeia	I don't want to pee	an0m&?0pi0	ad0wan0pi.	_def.	165	T			NN1	PNP VDB XX0 VVI TO0		2.65	0.265		
305	opal	hope I'll	.5p0l	h5p1l		165	T			NN1	VVB PNP VMB		0.27	0.054		
306	orgy	Porgy	.OrJI	pOrgI	title Progy and Bess	166	ne			NN1	NP0	Q+	0.68	0.136		
307	ovary	oh, very	5v0rI	5vErI		167	T			NN1	ITJ AV0		0.23	0.046	e	
308	paranoia	pair annoy you	p&r0n30	p&r0n30		169	T	homophonic		NN1	NN1 VVB PNP		0	0		
309	pasturage	past your age	p&sC0rIJ	p&sC0r4J	coll. past your age	171	T			NN1	PRP DPS NN1	I	0.28	0.035		
310	pistol	pissed till	pIst0l	pIstIl		176	T			NN1	VVD PRP		0	0	e	
311	pith	piss	pIT	pIs		177	ne		lisp	NN1	VVI		0.11	0.036666665		
312	pliers	players	pl10rz	pl40rz	_def.	178	ne			NN2	NN1		0.27	0.045		
313	possum	for some	pa.s0m	f0rs^m	idiom play possum	180	T			NN1	PRP AT0	I+	0.76	0.12666667	e	
314	psalmody	somebody	sam.0dI	s^mb0dI		184	T			NN1	NN1		0.42	0.06		
315	punishment	pun is meant	p^nISm0nt	p^nIzmEnt		185	T			NN1	NN1 VBB AV0		0.54	0.06		
316	pyorrhea	Piraeus	p10ri0.	p1.ri0s	_def.	186	ne			NN1	NP0		0.14	0.02		
317	qualm	calm	kwam	k.am	idiom the calm before the storm	188	B			NN1	NN1	E+	0.11	0.0275		
318	quondam	one damn	kwand0m	.w^nd&m	idiom one damn thing after another	189	T			NN1	AT0 AVO	I+	0.55	0.07857143		
319	are rapier	are happier	0r.4pi0r	0rh&pi0r		192	T			NN1	VBB AV0		0.34	0.0425		
320	reaper	ripper	rip0r	rIp0r	title Jack the Ripper	192	B			NN1	NN1	Q+	0.1	0.02		
321	rectangle	wrecked ankle	rEkt&Ng0l	rEkt&Nk0l	_def.	193	ne			NN1	AJ0 NN1		0.1	0.011111111	e	
322	reefers	refers	rif0rz	rIf9rz	_def.	194	T			NN2	VBZ		0.39	0.065	e	
323	reeling	reading	rilIN	ridIN	coll. reading and writing	257	B	see writhe		NN1	NN1	B	0.34	0.068		
324	refugee	referee	rEfjUJi	rEf.0ri	_def.	194	ne			NN1	NN1		0.77	0.11		
325	ribaldry	rivalry	rIb0ldrI	rIv0l.rI	coll. sibling rivalry	197	B			NN1	NN1	R	0.29	0.03625		
326	ricochet	rickshaw	rIk0S4	rIk.SO	_def.	197	ne			NN1	NN1		0.33	0.055		+1
327	rust	rushed	r^st	r^St		119	T	see indigo		NN1	VVN		0.17	0.0425		
328	safari	so far	s0farI	s5far.	idiom so far so good	201	T			NN1	AVO AVO	E+	0.35	0.058333334		
329	salmon	summon	sam0n	s^m0n	idiom summon one's strength	202	T			NN1	VVI	I+	0.21	0.042		
330	scrawl	call	skrOl	.k.Ol	title the call of the wild (London)	204	B			NN1	NN1	Q+	0.11	0.022		
331	script	stripped	skrIpt	strIpt		205	T			NN1	VVN		0.14	0.023333333		
332	sediment	said he meant	sEd0m0nt	sEdimEnt	_def.	205	ne			NN1	VVD PNP VVD		0.49	0.06125		
333	semaphore	same before	sEm.0fOr	s4mb0fOr	_def.	206	ne			NN1	PN1 AV0		0.49	0.06125		
334	seminary	cemetary	sEm0n7rI	sEm0t7rI	_def.	206	ne			NN1	NN1		0.34	0.0425		
335	senator	centaur	sEn0t0r	sEn.tOr	_def.	206	ne			NN1	NN1		0.29	0.04142857		+1
336	shasta	she has to	S.&st0	Si&st0		208	T			NN1	PNP VMZ TO0		0.29	0.048333332		-1
337	shiver	she ever	SI.v0r	SiEv0r		209	T			NN1	PNP AV0		1.15	0.19166666	e	
338	shofar	so far	S5f0r	s5far		209	T			NN1	AV0		0.35	0.07		
339	short	shore	SOrt	SOr.	coll. ship-to-shore radio	210	B	only morpheme		NN1	NN1	R	0.07	0.0175		
340	short	shot	SOrt	Sa.t	quote the shot heard round the world	210	B			NN1	NN1	Q+	0.28	0.07		
341	shotgun	shogun	Satg^n	S5.g^n		210	ne			NN1	NN1		0.32	0.053333335		
342	simile	smile	sIm0li	s.m1l.	_def.	211	ne			NN1	NN1		0.43	0.071666665		+2
343	smelt	smelled	smElt	smEld		213	T			NN1	VVD		0.11	0.022		
344	soda	so the	s5d0	s5D0		214	T			NN1	CJC AT0		0.05	0.0125		
345	Sodom	saw them	sad0m	sOD0m		214	T			NN1	VVD PNP		0.29	0.058		
346	soiree	sorry	swar4	s.arI	proverb love is never having to say you're sorry	215	T			NN1	AJ0	S+	0.39	0.078	l	
347	sophomore	sell for more	sa.f0mOr	sElf0mOr	idiom sell for more	215	T			NN1	VVB PNP AJC	I+	0.35	0.04375		
348	speculator	escalator	spE.kj0l4?0r	..Esk.0l4?0r	_def.	216	ne	this could be Red plus Lin		NN1	NN1		0.44	0.036666665		
349	stern	stone	st9rn	st5.n	coll. Rolling Stones	219	ne			NN1	NN1	R	0.32	0.064		
350	strumpet	strum pet	str^mpIt	str^mpIt	_def.	221	ne	homophonic		NN1	NN0 NN1		0	0		
351	summons	someone's	s^m.0nz	s^mw^nz		222	T			NN2	PNP POS		0.4	0.057142857		
352	T.b.	to be	tibi	t0bi	quote to be or not to be	54	B			NN1	TO0 VBB	Q+	0.24	0.06		
353	tantrum	tandem	t&ntr0m	t&nd.0m	_def.	226	B			NN1	NN1		0.14	0.02		

354	termite	their might	t9rm1t	D7rm1t	coll. with all one's might	229	ne			NN1	DPS NN1	I	0.47	0.07833333	e	
355	throat	throw it	Tr5.t	Tr5It	idiom throw it out	230	T			NN1	VVB PNP	I+	0.11	0.022		-1
356	ticket	thicket	tIk0t	TIk0t		231	ne		lisp?	NN1	NN1		0.18	0.036		
357	toboggan	to bargain	t0ba.g0n	t0barg0n	_def.	232	ne			NN1	TO0 VVI		0.04	0.005		
358	tog	tug	tag	t^g	idiom tug of war	233	ne			NN1	NN1	I+	0.21	0.07		
359	togs	dogs	tagz	dagz	coll. mad dogs	232	ne			NN2	NN2	C	0.11	0.0275		
360	bit tonsure	bit unsure	bItanS0r	bIt^nS8r	idiom a bit unsure of oneself	233	T			NN1	AV0 AJ0	E+	0.51	0.06375		
361	trial	try all	tr10l	tr1Ol		235	T			NN1	VVI AV0		0.25	0.05	e	
362	trout	out	tr2t	..2t		236	T			NN1	AVP		0.1	0.025		
363	urchin	urgin'	9rCIn	9rJIn	coll. urge someone on	240	T			NN1	VVG	P	0.2	0.04		
364	urination	you're a nation	jUr0n4S0n	j8r0n4S0n	idiom you're a nation of ...	240	T			NN1	PNP VBB AT0 NN1	Q+	0.29	0.032222223		
365	V	we	vi	wi	idiom where do we go from here	242	T			NN1	NN1	I+	0.08	0.04		
366	van	when	v&n	wEn		242	T	syntactically		NN1	AVQ		0.27	0.09		
367	vanilla	Manilla	v0nIl0	m0nIl0	_def.	242	ne			NN1	NP0		1	0.16666667		
368	vast	last	v&st	l&st	title the last of the Mohican	243	T			NN1	NN1	Q+	0.44	0.11		
369	veal	feel	vil	fil	idiom only as young as you feel	243	T			NN1	VVB	E+	0.17	0.05666667		
370	vermilion	million	v0rmIlj0n	...mIlj0n		118	T	see indigo		NN1	NN1		0.41	0.045555554		+1
371	version	virgin	v9rZ0n	v9rJ0n		244	ne			NN1	NN1		0.31	0.051666666		
372	viscount	why count	v1k2nt	w1k2nt		246	T			NN1	AVQ VVB		0.08	0.013333334		
373	vodka	what ca(n)	vadk0	watk0		199	T	see rum		NN1	AVQ VMB		0.17	0.034		
374	wafer	wait for	w4.f0r	w4tf0r	idom swallow one and wait for results	248	T			NN1	VVB AVP	I	0.07	0.011666667		
375	whittler	Hitler	wItl0r	hItl0r	coll. Heil Hitler	252	ne			NN1	NP0	F	0.38	0.06333333		
376	wiener	when are	win0r	wEn0r		253	T			NN1	AVQ VBB		0.28	0.056		
377	wimble	nimble	wImb0l	nImb0l	quote Jack be nimble, Jack be quick ...	253	T			NN1	AJ0	Q+	0.52	0.086666666		
378	wind	when d(o)	wInd	wEnd	idiom run out of wind	254	T			NN1	AVQ VDB	E+	0.2	0.05		
379	wretched	Richard	rECI.d	rIC0rd	coll. Richard the Lionhearted	256	B			NN1	NP0	I	0.26	0.043333333		
380	writhing	writing	r1DIN	r1?IN	coll. reading and writing	257	B			NN1	NN1	B	1	0.2		
381	yacht	he ought	j.at	hiOt		258	T			NN1	PNP VMB		0.97	0.2425		-1
382	yarn	you're on	jar.n	j8ran	idiom you're on the wrong track	259	T			NN1	PNP VBB AVP	E+	0.95	0.19		-1
383	yelp	help	jElp	hElp	coll. helpless	259	B	just morpheme		NN1	NN1	R	0.44	0.11		
384	yokel	yodel	j5k0l	j5d0l	_def.	260	ne			NN1	VVI		0.34	0.068		
385	zend	send	zEnd	sEnd		261	T	what?		VVN	VVN		0.07	0.0175		
386	zinc	sink	zINk	sINk	_def.	262	ne			NN1	NN1		0.07	0.0175		
387	zombie	some be	zambi	s^mbi		262	T			NN1	PNP VBB		0.28	0.056		
388	zounds	sounds	z2ndz	s2ndz	_def.	262	ne	what?		NN2	NN2		0.07	0.014		
389	poplar	popular	pOp..l0r	pOpj0l0r		17	T	see aspen		NN1	AJ0		0	0		-1
390	alder	all the	Old0r	0ld0.	coll. all the time	17	T	see aspen		NN1	AV0 AT0	I	0.3	0.06		
391	nutmeg	not make	natmEg	nOtm4k		54	T	see consecrated		NN1	XX0 VVB		0	0	e	
392	camphor	came for	k&mf0r	k4mfOr		54	T	see consecrated		NN1	VVD AVP		0	0		
393	sulphur	sell for	salf0r	sElfOr		54	T	see consecrated		NN1	VVB AVP		0	0		
394	cinnamon	seen a man	sIn0m0n	sin0m&n		54	T	see consecrated		NN1	VVN AT0 NN1		0	0	e	
395	ammonia	I'm only a	&m5n.6	1m5nI6		54	T	see consecrated		NN1	PNP VBB AV0 AT0		0	0		
396	hair	heir	h&r	.ir	coll. heir apparent	106	ne			NN1	NN1	R	0	0		
397	pun	pen	pan	pEn	idiom put pen to paper	124	B	see jest		NN1	NN1	E+	0	0		
398	beer	be or	bi.r	biOr		199	T	see rum		NN1	AV0 CJC		0	0		-1
399	water	what do	wOd0r	wOt0.		199	T	see rum		NN1	AVQ VDB		0	0		
400	wine	why	w1n	w1.		199	T	see rum		NN1	AVQ		0	0		+1
401	martini	are tiny	martini	.art1nI		199	T	see rum		NN1	VBB AV0		0	0		
402	chianti	can't you	kj&nt.I	k.&ntjU		199	T	see rum		NN1	VM0 XX0 PNP		0	0	l	
403	sop horrific	soporific	saph0rIfIk	sap.0rIfIk		215	B			NN1 AV0	NN1		0	0		
404	tryer	cryer	tr10r	kr10r	coll. town cryer	236	B			NN1	NN1	R	0.16	0.032		
405	abbess	abscess	&b.0s	&bs0s		5	B			NN1	NN1		0.1	0.02		
406	admonition	ammunition	&dm.0nIS0n	&.mjunIS0n	title praise the lord and pass the ammunition	8	B			NN1	NN1	Q+	0.48	0.048		
407	allusion	illusion	0luZ0n	IluZ0n	coll. optical illusion	11	B			NN1	NN1	C	0	0		
408	ammonia	(th)em on yet	0m5nj0.	0manjEt		12	B	should it be them on, ya?		NN1	NN1 PNP AV0		0.57	0.08142857		
409	anapest	and a pest	&n0pEst	0n0pEst	idiom become a pest	13	B			NN1	CJC AT0 NN1	I+	0.23	0.032857142	e	
410	anti-skeptic	antiseptic	&n?0skEptIk	&n?0s.EptIk	coll. antispetic	212	ne	only morpheme		NN1	NN1	R	0.1	0.009090909		
411	arrival	survival	.0r.1v0l	s0rv1v0l	idiom survival of the fittest	16	B	much distance, but idiom		NN1	NN1	Q+	0.27	0.03375		
412	asp	ass	&sp	&s.	idiom make an ass of oneself	17	B			NN1	NN1	E+	0.16	0.053333335		
413	baboons	bassoons	b0bunz	b0sunz		245	B	see vile		NN2	NN2		0.66	0.11		
414	barbiturate	bar bitch you ate	barbIC0rIt	barbIC0.Et		23	B	oh, what a setup		NN1	NN0 NN1 PNP		0.23	0.023		
415	bauble	bubble	bOb0l	b^b0l		25	B	what?		NN1	NN1		0.28	0.056		

416	bay	day	b4	d4	proverb every dog has his day	25	B			NN1	NN1	S+	0.21	0.105	
417	beach	peach	biC	piC		25	B	or 'bitch' [bIC], 3-way pun		NN1	NN1		0.14	0.046666667	
418	beans	beings	binz	biNz	coll. human beings	26	B			NN2	NN2	C	0.14	0.035	
419	beard	bard	b6rd	bard	coll. Bard of Avon	26	B			NN1	NN1	I	0.3	0.075	
420	biggie	buggy	bIgI	b^gI	coll. dune buggie	29	B			NN1	NN1	R	0.27	0.0675	
421	blond	blind	bland	bl1nd	coll. Venetian blind	31	B			NN1	NN1	C	0.23	0.046	
422	braid	bread	br4d	brEd	idiom cast one's bread upon the water	34	B	what?		NN1	NN1	I+	0.26	0.065	
423	braziers	brassiers	braZ0rz	br0z6rz		34	B			NN2	NN2		0.91	0.13	
424	breath	breadth	brE.T	brEdT	idiom by a hair's breadth	34	B			NN1	NN1	I+	0.09	0.018	
425	bream	dream	brim	drim	idiom dream the impossible dream	34	B	adnominatio		NN1	NN1	E+	0.21	0.0525	
426	breaths	breasts	brE.Ts	brEsts		34	B			NN2	NN2		0.3	0.05	
427	brogan	broken	br5g0n	br5k0n		36	B			NN1	AJ0		0.1	0.016666668	
428	bruise	blues	bruz	bluz		36	B		Chinese	NN1	NN1		0.13	0.0325	
429	bungle	bundle	b^Ng0l	b^nd0l	idiom bundle of joy	38	B			NN1	NN1	I+	0.63	0.105	
430	butcher	bitch or	bUC0r	bIC0r		39	T			NN1	NN1 CJC		0.3	0.06	
431	can	man	k&n	m&n	title a man for all seasons	41	B			NN1	NN1	Q+	0.66	0.22	
432	carp	cart	karp	kart	idiom put the cart before the horse	43	B			NN1	NN1	E+	0.18	0.045	
433	centimeter	centipede	sEn?Imi?0r	sEn?Ipid..		45	B			NN1	NN1		0.91	0.091	+1
434	chimps	chips	CImps	CI.ps	coll. fish and chips	87	B	see finch		NN2	NN2	B	0.14	0.028	
435	chintzes	chances	CIntsIz	C&n.sIz	idiom take your chances	48	B			NN2	NN2	I+	0.35	0.05	
436	clown	crown	kl2n	kr2n	quote Jack fell down and broke his crown ...	50	B			NN1	NN1	Q+	0.11	0.0275	
437	coal	cold	k5l.	k5ld	coll. a bad cold	50	B			NN1	NN1	C	0.09	0.0225	
438	colt	cold	k5lt	k5ld	idiom feed a cold and starve a fever	52	B			NN1	NN1	E+	0.11	0.0275	
439	condiment	complement	kand.Im0nt	kamplIm0nt		54	B	see consecrated		NN1	NN1		1.22	0.122	
440	congestion	question	k0nJEsC0n	k..wEsC0n	quote to be or not to be ...	54	B			NN1	NN1	Q+	0.8	0.08888889	+1
441	cope	hope	k5p	h5p	proverb where there's life, there's hope	55	B			NN1	NN1	S+	0.38	0.12666667	
442	corsets	charges	kOrsIts	CarJ0z	coll. charges without foundation	56	B	Ws had corsets-causes		NN2	NN2	I	2.78	0.46333334	
443	cortege	corsage	kOrtEZ	kOrsaZ		56	B			NN1	NN1		0.73	0.12166667	
444	crab	grab	kr&b	gr&b		57	B			NN1	NN1		0.11	0.0275	
445	critic	cricket	krI?Ik	krIkIt	title cricket on the hearth (Dickens)	58	B			NN1	NN1	Q+	0.8	0.13333334	
446	cruisers	choosers	kruz0rz	C.uz0rz	idiom beggars can't be choosers	59	B			NN2	NN2	S+	0.55	0.07857143	
447	crullers	crawlers	kr^l0rz	krOl0rz	idiom night crawlers	59	B			NN2	NN2	R	0.28	0.04	
448	cub	cud	k^b	k^d	_def.	60	B			NN1	NN1		0.21	0.07	
449	curries	courage	k9rIz	k9rIJ	idiom have the courage of your conviction	60	B			NN2	NN1	I+	1	0.2	
450	curse	course	k9rs	kOrs	idiom let nature take its course	61	B			NN1	NN1	E+	0.29	0.0725	
451	delight	the light	dIl1t	D0l1t	coll. lower the light	64	B			NN1	AT0 NN1	I	0.06	0.012	
452	dependants	the pendants	dIpEnd0nts	D0pEnd0nts		65	B			NN2	AT0 NN2		0.06	0.006	
453	din	den	dIn	dEn	idiom den of iniquity	67	B			NN1	NN1	E+	0.2	0.06666667	
454	dinosaur	diner sore	d1n0.sOr	dIn0rsOr		67	B			NN1	NN1 AJ0		0.32	0.04	
455	doe	do	d5	du		68	B	what?		NN1	NN1		0.27	0.135	
456	dog	log	dag	lag	idiom throw a log on the fire	68	B	fire-log eskimo-sleddog		NN1	NN1	E+	0.44	0.14666666	
457	dowel	towel	d20l	t20l	idiom throw in the towel	69	B			NN1	NN1	E+	0.09	0.0225	
458	dram	damn	dr&m	d.&m	coll. give a damn	69	B			NN1	NN1	T	0.03	0.0075	
459	drool	drawl	drul	drOl	coll. Southern drawl	70	B			NN1	NN1	C	0.29	0.0725	
460	dung	done	d^N	d^n	title when the works all done this fall	72	B			NN1	VVN	Q+	0.19	0.06333333	
461	earl	oil	9rl	3.l	idiom pour oil on troubled water	73	B			NN1	NN1	E+	0.33	0.11	
462	ears	years	.6rz	j6rz	idiom give the best years of one's life	73	B			NN2	NN2	I+	0.09	0.0225	
463	eaves	heaves	.ivz	hivz	coll. have the heaves	74	B			NN2	NN2	T	0.07	0.0175	
464	election	erection	IlEkS0n	IrEkS0n		75	B		Chinese	NN1	NN1		0.11	0.015714286	
465	elephants	eloquence	El0f.0ns	El0kw0ns	title well-timed silence... (M. F. Tupper)	76	B			NN2	NN1	Q+	0.57	0.07125	
466	ember	amber	Emb0r	&mb0r	title Forever Amber (movie)	76	B			NN1	NN1	Q+	0.2	0.04	
467	epistle	a pistol	ipIs.0l	0pIst0l		78	B			NN1	AT0 NN1		0.31	0.044285715	
468	etch	itch	EC	IC	idiom scratch the itch	79	B			NN1	NN1	E+	0.21	0.105	
469	ether	either	iT0r	iD0r		80	B	either = pronoun		NN1	NN1		0.23	0.0575	
470	eye-deer	idea	1d6r	1d6.	coll. interesting idea	81	B			NN1	NN1	C	0.03	0.0075	
471	fart	part	fart	part	idiom part and parcel	84	B			NN1	NN1	B+	0.23	0.0575	
472	faucet	facet	fOsIt	f&sIt	coll. facets of emotion	84	B			NN1	NN1	I	0.29	0.058	
473	fervor	further	f9rv0r	f9rD0r	idiom so far and no further	86	B			NN1	AV0	E+	0.38	0.06333333	
474	fife	wife	f1f	w1f		86	B			NN1	NN1		0.34	0.11333334	
475	finch	fish	fInC	fI.S	coll. fish and chips	87	B			NN1	NN1	B	0.29	0.0725	
476	fissure	figure	fIS.0r	fIgj0r	quote after the figure of man (Isa 44:13)	88	B			NN1	NN1	Q+	1.09	0.18166667	
477	flies	(ha)ve eyes	fl1z	v.1z	title I only have eyes for you	90	B	1 syll lost		NN2	VHB NN1	Q+	0.2	0.05	

478	focus	hocus	f5k0s	h5k0s	coll. hocus-pocus'	90	B	non-word		NN1	NN1	R	0.52	0.104		
479	foible	fable	f3b0l	f4b0l		90	B			NN1	NN1		0.29	0.058		
480	fort	fourth	fOrt	fOrT		88	B	see fit	AAVE	NN1	NN1		0.18	0.045		
481	fowl	fall	f2l	fOl	idiom fall of the law	92	B			NN1	NN1	E+	0.29	0.096666664		
482	fowl	fool	f2l	ful	proverb there's no fool like an old fool	92	B			NN1	NN1	S+	0.29	0.096666664		
483	fronds	friends	frandz	frEndz	idiom visit a sick friend	93	B			NN2	NN2	I+	0.27	0.045		
484	futility	utility	fjutIl0tI	.jutIl0tI	coll. public utility	94	B			NN1	NN1	C	0.27	0.03		
485	game	gay	g4m	g4.		95	B			NN1	AJ0		0.14	0.046666667		
486	gargoyles	gargles	garg3lz	garg0lz		96	B			NN2	NN2		0.3	0.042857144		
487	ghoul	fool	gul	ful	proverb there's no fool like an old fool	97	B	see fowl		NN1	NN1	S+	0.52	0.17333333		
488	glance	dance	gl&ns	d.&ns	idiom may I have the next dance	98	B			NN1	NN1	I+	0.5	0.1		
489	gloom	room	glum	.rum	coll. room for improvement	98	B			NN1	NN1	I	0.29	0.0725		
490	God	gard	ga.d	gard	New England	99	B			NN1	NN1		0.04	0.01		
491	gourd	word	gOrd	w9rd	quote ... never heard a discouraging word ...	100	B	what?		NN1	NN1	Q+	0.95	0.2375		
492	gripe	grape	gr1p	gr4p		102	B	what?		NN1	NN1		0.27	0.0675		
493	groin	coin	gr3n	k.3n	idiom toss a coin in the fountain	102	B			NN1	NN1	E+	0.13	0.0325		
494	grouse	gross	gr2s	gr5s	proverb ... always tell a girl is game ...	103	B	what?		NN1	AJ0	S+	0.29	0.0725		
495	hake	cake	h4k	k4k	quote let them eat cake (M. Antoinette)	106	B			NN1	NN1	Q+	0.34	0.11333334		
496	hash	haste	h&S.	h4st	proverb more speed, less haste	108	B			NN1	NN1	S+	0.43	0.1075		
497	heap	sheep	hip	Sip	proverb it doesn't take a smart girl ...	109	B			NN1	NN1	S+	0.66	0.22		
498	heresy	hair I see	hErIsI	h7r1si	idiom hair in the soup	110	B			NN1	NN1 PNP VVB	E+	0.71	0.11833333	e	
499	hip	lip	hIp	lIp	idiom stiff upper lip	111	B			NN1	NN1	C+	1	0.33333334		
500	horses	hoses	hOrsIz	h5.zIz	idiom firemen play their hoses	113	B			NN2	NN2	E+	0.38	0.06333333		
501	hull	hell	h^l	hEl	coll. hell of a place	114	B	WS had hull-whole		NN1	NN1	I	0.26	0.086666666		
502	humor	human	hjum0r	hjum0n	proverb to err is human	114	B			NN1	NN1	S+	0.31	0.051666666		
503	hurter	hitter	h9r?0r	hI.?0r		114	B	what?		NN1	NN1		0.33	0.055		
504	icicle	bicycle	.1sIk0l	b1sIk0l	_def.	116	B			NN1	NN1		0.21	0.03		
505	ilk	milk	.Ilk	mIlk	idiom milk of human kindness	117	B			NN1	NN1	E+	0.19	0.0475		
506	impotentate	potentate	Imp5?0nt4t	..p5?0nt4t		118	T	non-word in P, syllable loss		NN1	NN1		0.24	0.024		+1
507	interdict	into Dick	In?0rdIkt	In?0.dIk.		121	T			NN1	PRP NN1		0.1	0.011111111	e	
508	invoices	voices	Inv3sIz	..v3sIz	idiom hear strange voices	121	B			NN2	NN2	I+	0.2	0.028571429		+1
509	irk	work	.9rk	w9rk	proverb the devil finds work for idle hands	122	B			NN1	NN1	S+	0.13	0.0325		
510	jackal	Jack hell	J&k.0l	J&khEl	coll. give someone hell	123	B			NN1	NP0 NN1	I	0.3	0.05		
511	jag	rag	J&g	r&g	idiom a rag, a bone, and a hank of hair	123	B			NN1	NN1	E+	0.38	0.12666667		
512	jilting	tilting	JIl?IN	tIl?IN		125	B	jolting?		NN1	NN1		1	0.16666667		
513	jugs	bugs	J^gz	b^gz	idiom get the bugs out	126	B			NN2	NN2	E+	1	0.25		
514	kilt	killed	kIlt	kIld		129	B			NN1	VVN		0.11	0.0275		
515	kin	skin	.kIn	skIn	quote skin you love to touch (Woodb. Soap)	129	B			NN1	NN1	Q+	0.1	0.025		
516	kingdom	king dumb	kINd0m	kINd^m		130	B			NN1	NN1 AJ0		0.27	0.045		
517	kitty	pity	kI?I	pI?I	idiom more's the pity	130	B			NN1	NN1	E+	0.23	0.0575		
518	largesse	large ass	larJEs	larJ&s		134	B			NN1	AJ0 NN1		0.2	0.033333335	l	
519	limb	lamb	lIm	l&m	title Mary had a little lamb	138	B			NN1	NN1	Q+	0.28	0.093333334		
520	linden	linen	lInd0n	lIn.0n	proverb wash one's linen in public (chiefly BE)	138	B			NN1	NN1	S+	0.11	0.018333333		
521	lisp	Liszt	lIsp	lIst		139	B			NN1	NP0		0.18	0.045		
522	locomotive	local motive	l5k0.m5?Iv	l5k0lm5?Iv		141	B			NN1	AJ0 NN1		0.08	0.008		
523	lodger	ledger	laJ0r	lEJ0r	coll. dusty legder	141	B			NN1	NN1	C	0.27	0.054		
524	loon	moon	lun	mun	coll. full moon	142	B			NN1	NN1	C	0.38	0.12666667		
525	lust	lost	l^st	last	proverb it is better to have loved and lost ...	143	B	two diff synt. contexts		NN1	VVN	S+	0.26	0.065		
526	medication	meditation	mEdIk4S0n	mEdIt4S0n	coll. transcendental meditation	148	B			NN1	NN1	C	0.14	0.015555556		
527	menopause	men who pause	mEn.0pOz	mEnhupOz		148	B			NN1	NN1 DTQ VVB		0.35	0.04375		
528	millstone	milestone	mIlst5n	m1lst5n	idiom a milestone on the road to ...	150	B			NN1	NN1	E+	0.28	0.04		
529	mirage	marriage	mIraZ	m&rlJ	idiom marriage of convenience	150	B			NN1	NN1	I+	0.87	0.174	l	-1
530	moan	mowin'	m5.n	m5In	idiom mowing of the grass	151	B			NN1	NN1	I+	0.11	0.0275		-1
531	moats	most	m5ts	m5.s	idiom get the most for one's money	151	B			NN2	NN1	I+	0.07	0.0175		
532	monsoon	man soon	mansun	m&nsun		152	B			NN1	NN1 AV0		0.18	0.03	l	
533	mosque	moss	mask	mas.	proverb a rolling stone gathers no moss	153	B			NN1	NN1	S+	0.1	0.025		
534	muscle	missile	m^s0l	mIs0l	coll. guided missile	154	B			NN1	NN1	C	0.29	0.058		
535	narc	knock	nark	na.k	idiom knock on the door	157	B			NN1	NN1	E+	0.03	0.0075		
536	niece	knees	nis	niz		159	B			NN1	NN2		0.07	0.023333333		
537	night	light	n1t	l1t	quote let there be light ...	159	B			NN1	NN1	Q+	0.23	0.07666667		
538	node	note	n5d	n5t	coll. keep notes	160	B			NN1	NN1	I	0.09	0.03		
539	none	one	n^n	w^n	idiom one in a million	160	B			NN1	NN1	E+	0.52	0.17333333		

540	Norse	horse	nOrs	hOrs	idiom a horse of a different color	161	B		NN1	NN1	E+	0.52	0.13	
541	O-van	oven	5v&n	5v0n		242	B	non-word	NN1	NN1		0.23	0.0575	
542	parakeets	pair of tits	p&r0kits	p7r0tIts	coll. pair of tits	169	B		NN2	NN1 PRP NN1	I	0.5	0.0625	
543	parrots	Paris	p&r0ts	p&r0.s	title I love Paris in the springtime	170	B	WS had parrots-carrots	NN2	NP0	Q+	0.07	0.011666667	
544	parrots	parents	p&r0.ts	p&r0nts	quote love thy parents	170	B		NN2	NN2	Q+	0.1	0.014285714	
545	paunch	punch	pOnC	p^nC	idiom pull one's punches	171	B		NN1	NN1	E+	0.28	0.07	
546	pest	post	pEst	p5st	idiom desert one's post	174	B		NN1	NN1	E+	0.29	0.0725	
547	phrase	phase	fr4z	f.4z	idiom a phase one's passing through	175	B		NN1	NN1	E+	0.03	0.0075	
548	piety	pie & tea	p10.tI	p10nti	coll. pie and tea	176	B		NN1	NN1 CJC NN1	B	0.25	0.041666668	
549	pill	will	pIl	wIl	proverb where there's a will, there's a way	176	B		NN1	NN1	S+	1	0.33333334	
550	pool	fool	pul	ful	proverb there's no fool like an old fool	179	B		NN1	NN1	S+	0.21	0.07	
551	potpourri	popery	p5puri	p5p0rI		180	B		NN1	NN1		0.36	0.06	
552	preposition	proposition	prEp0zIS0n	prap0zIS0n	coll. make propositions	181	B		NN1	NN1	I	0.29	0.029	
553	prism	prison	prIz0m	prIz0n	proverb iron bars do not a prism make	182	B		NN1	NN1	S+	0.11	0.018333333	
554	pun	gun	p^n	g^n	idiom point a gun	184	B		NN1	NN1	I+	0.66	0.22	
555	pun	bun	p^n	b^n	idiom throw hot, cross buns	184	B		NN1	NN1	E+	0.16	0.053333335	
556	pyromaniac	pie maniac	p1r5m4ni0k	p1..m4ni0k		187	B		NN1	NN0 NN1		0.27	0.027	+1
557	quake	cake	kw4k	k.4k	proverb you can't have your cake and eat it	188	B		NN1	NN1	S+	0.11	0.0275	
558	quarrel	barrel	kwOr0l	b.&r0l	title roll out the barrel	188	B		NN1	NN1	Q+	0.92	0.15333334	
559	quench	wench	kwEnC	.wEnC	coll. lusty wench	188	B		NN1	NN1	C	0.1	0.02	
560	query	queer	kwirI	kwir.		188	B		NN1	NN1		0.39	0.078	+1
561	quill	will	kwIl	.wIl	proverb where there's a will there's a way	189	B		NN1	NN1	S+	0.1	0.025	
562	quirk	work	kw9rk	.w9rk	idiom all work and no pay	189	B		NN1	NN1	S+	0.1	0.02	
563	quiver	giver	kwIv0r	g.Iv0r	quote the Lord loveth a cheerful giver	189	B		NN1	NN1	Q+	0.22	0.036666665	
564	rabbit	habit	r&bIt	h&bIt	idiom kick the habit	190	B		NN1	NN1	T+	0.66	0.132	
565	Rachmaninov	man enough	r&km&nInOf	...m&nIn^f	idiom not man enough	190	B		NN1	NN1 AV0	E+	0.93	0.093	+1
566	reef	wreath	rif	riT	coll. coral reef	193	B		NN1	NN1	C	0.24	0.08	
567	resolution	revolution	rEz0luS0n	rEv0luS0n		195	B		NN1	NN1		0.52	0.057777777	
568	rib-off	ripoff	rIbaf	rIpaf	idiom that's a rip-off	196	B		NN1	NN1	P+	0.14	0.028	
569	robot	row boat	r5bat	r5b5t		198	B		NN1	NN0 NN1		0.25	0.05	
570	rooks	books	rUks	bUks		198	B	spoonerism	NN2	NN2		0.66	0.165	
571	rubble	trouble	.r^b0l	tr^b0l	title nobody knows the trouble I've seen	199	B		NN1	NN1	Q+	0.07	0.011666667	
572	rupture	rapture	r^pC0r	r&pC0r	coll. sheer rapture	200	B		NN1	NN1	C	0.28	0.046666667	
573	ruse	rose	ruz	r5z	quote a rose is a rose is a rose	200	B		NN1	NN1	Q+	0.27	0.09	
574	rusk	risk	r^sk	rIsk	idiom run the risk	200	B		NN1	NN1	I+	0.29	0.0725	
575	S	ass	Es	&s	idiom make an ass of oneself	201	B		NN1	NN1	E+	0.2	0.1	
576	salivation	salvation	s&lIv4S0n	s&l.v4S0n		202	B		NN1	NN1		0.1	0.011111111	+1
577	sarong	(what')s wrong	s0rON	s.rON	idiom what's wrong	203	ne		NN1	VBB AVO	I+	0.04	0.008	
578	scar	star	skar	star	title stars and stripes forever	203	B		NN1	NN1	Q+	0.14	0.035	
579	scarlets	harlots	skarlIts	.harl0ts	coll. film starlet	203	B		NN2	NN2	R	0.47	0.05875	
580	scherzo	shirt so	sk7rts5	S.9rts5		204	B		NN1	NN1 AV0		0.57	0.08142857	
581	screw	shrew	skru	S.ru	title the taming of the shrew	204	B		NN1	NN1	Q+	0.27	0.0675	
582	scrod	screwed	skrad	skrud	coll. get screwed	205	B		NN1	VVN	I	0.3	0.06	
583	sherry	cherry	SErI	CErI	idiom cut wine	209	B		NN1	NN1	I+	0.15	0.0375	
584	shoat	shot	S5t	Sat	quote fired the shot heard round the world	209	B		NN1	NN1	Q+	0.27	0.09	
585	sin	gin	sIn	JIn	coll. gin and tonic	211	B		NN1	NN1	B	1	0.33333334	
586	sing	thing	sIN	TIN	idiom it don't mean a thing	211	B		NN1	NN1	E+	0.25	0.083333336	
587	sitar	sitter	sI?ar	sI?0r	coll. baby sitter	211	B		NN1	NN1	R	0.24	0.048	
588	snake	shake	sn4k	S.4k	coll. fair shake	214	B		NN1	NN1	C	0.27	0.0675	
589	snore	snare	snOr	sn7r	coll. a snare and a delusion	214	B		NN1	NN1	B	0.29	0.0725	
590	spies	eyes	sp1z	..1z	title can't keep my eyes off of you	218	B	syllable to be important	NN2	NN2	Q+	0.24	0.06	
591	sports	spots	spOrts	spa.ts	proverb a leopard can't change its spots	217	B		NN1	NN2	S+	0.28	0.046666667	
592	squid	kid	skwId	.k.Id	idiom a crazy mixed-up kid	218	B		NN1	NN1	E+	0.19	0.038	
593	stock	stuck	stak	st^k		220	T		NN1	AJ0		0.21	0.0525	
594	swish	wish	swIS	.wIS	idiom grant your wish	224	B		NN1	NN1	I+	0.08	0.02	
595	talc	talk	t&lk	tO.k	idiom talk of the town	226	B		NN1	NN1	E+	0.33	0.0825	
596	tan	man	t&n	m&n	idiom richest man in town	226	B		NN1	NN1	E+	1	0.33333334	
597	taste	paste	t4st	p4st	coll. scissors and paste	227	B		NN1	NN1	B	0.31	0.0775	
598	teams	times	timz	t1mz	proverb small minds can't handle great times	227	B		NN2	NN2	S+	0.28	0.07	
599	teat	to eat	t.it	t0it	idiom what's to eat	228	B		NN1	TO0 VVI	I+	0.07	0.0175	
600	tens	tents	tEnz	tEns	idiom fold one's tents and steal away	228	B		NN2	NN2	E+	0.07	0.0175	
601	thrall	thrill	TrOl	TrIl	idiom the thrill of a lifetime	230	B		NN1	NN1	I+	0.3	0.075	

602	tight	right	t1t	r1t	idiom defend one's rights	231	B			NN1	NN1	I+	0.52	0.17333333		
603	time	nine	t1m	n1n	proverb a stitch in time saves nine	214	B	see snitch		NN1	NN1	S+	0.77	0.25666666		
604	trillium	trillion	trIli0m	trIli0n	idiom make a million/trillion	235	B			NN1	NN1	I+	0.11	0.015714286		
605	tug	bug	t^g	b^g	idiom get the bugs out	237	B			NN1	NN1	E+	0.38	0.12666667		
606	typhoon	tycoon	t1fun	t1kun	coll. business tycoon	238	B			NN1	NN1	R	0.44	0.088		
607	umber	rumba	.^mb0r	r^mb0.	idiom dance a rumba	119	T	WS had umber-number		NN1	NN1	I+	0.07	0.011666667		
608	Uranus	your anus	jur4n0s	jOr4n0s		240	B			NN2	DPS NN1		0.29	0.04142857	e	
609	valet	valley	v&lIt	v&lI.	title how green was my valley	242	B			NN1	NN1	Q+	0.07	0.014		
610	veil	bail	v4l	b4l	quote old man river	243	B			NN1	NN1	Q+	0.44	0.14666666		
611	(a) viaduct	(to) buy a duck	v10d^kt	b10d^k.		244	B			NN1	VVI AT0 NN1		0.51	0.07285714		
612	vice	spice	.v1s	sp1s	idiom the spice of life	245	B			NN1	NN1	E+	0.76	0.19		
613	viles	violins	v1.l..z	v10lInz		245	B			NN2	NN2		0.28	0.04		-1
614	villain	willin'	vIIIn	wIIIn	idiom ready, willing, and able	245	B			NN1	AJ0	B+	0.08	0.016		
615	volley	folly	valI	falI	proverb war is folly	246	B			NN1	NN1	S+	0.17	0.0425		
616	volt	fault	v5lt	fOlt	idiom it wasn't my fault	247	B			NN1	NN1	I+	0.43	0.1075		
617	warm	worm	wOrm	w9rm	proverb the early bird gets the worm	249	B			NN1	NN1	S+	0.29	0.0725		
618	whore	tour	hOr	t8r	coll. tour de force	252	B			NN1	NN1	F	1.29	0.43		
619	wine	line	w1n	l1n	idiom line of least resistance	254	B			NN1	NN1	E+	0.25	0.083333336		
620	wiper	viper	w1p0r	v1p0r	idiom take a viper to one's bosom	254	B			NN1	NN1	E+	1	0.2		
621	writs	wits	rIts	wIts	idiom have one's writs about one	256	B			NN2	NN2	E+	0.21	0.0525		
622	youth	use	juT	jus	idiom what's the use?	261	B			NN1	NN1	E+	0.11	0.036666665		
623	zinnia	seein' you	zInj0	sinj0	idiom I'll be seeing you	262	B			NN1	VVG NN1	I+	0.22	0.044		
624	boa	bore	b50	bOr		7	B	overlooked by WS		NN1	NN1		0	0		
625	carps	cops	karps	kO.ps	coll. cops and robbers	43	B			NN2	NN2	B	0	0		
626	fry	fly	fr1	fl1	idiom fly in the ointment	94	B			NN1	NN1	E+	0	0		
627	hair	here	h&r	hir	idiom here today, gone tomorrow	106	B	elliptical		NN1	AV0	E+	0	0		
628	tequila	to kill ya	tEkil.0	t0kIlj0		199	B	see rum		NN1	TO0 VVI PNP		0	0		
629	sheep	cheap	Sip	Cip	proverb buy cheap, sell dear	209	B			NN1	NN1	S+	0	0		
630	sheik	shreak	S.ik	Srik	coll. terrified shreak	209	B			NN1	NN1	C	0	0		
631	strumpet	trumpet	str^mpIt	.tr^mpIt		245	B	see vile		NN1	NN1		0.08	0.01		
632	pastel	past her	p&stEl	p&st0r		118	T	see indigo, that's pushing it		AJ0	PRP PNP		0.37	0.061666667		
633	for heifer	forever	f0hEf0r	f0.Ev0r		109	B			PRP NN1	AV0		0.22	0.03142857		
634	up ill	uphill	^p.Il	^phIl		117	T	what?		PRP AJ0	AV0		0.07	0.014		
635	amorous	him or us	.&m0r0s	hIm0r0s	idiom it's them or us	12	B			AJ0	PNP CJC PNP	E+	0.33	0.047142856		
636	angora	angry	&NgOr0	&Ng.rI	idiom don't be angry	16	B	see artesian		AJ0	AJ0	I+	0.52	0.086666666	l	+1
637	bartered	buttered	bar?0rd	b^.?0rd	idiom which side one's bread is buttered on	25	B			AJ0	AJ0	E+	0.24	0.034285713		
638	black	back	bl&k	b.&k	idiom back right where one begun	119	B	see indigo		AJ0	AJ0	S+	0.06	0.015		
639	dilated	delighted	d1l4?Id	dIl1?Id	idiom delighted to meet you	67	B			AJ0	AJ0	I+	0.54	0.07714286		
640	draped	raped	dr4pt	.r4pt		69	B			AJ0	AJ0		0.11	0.022		
641	drugged	dragged	dr^gd	dr&gd		70	B			AJ0	AJ0		0.28	0.056		
642	emissary	necessary	.EmIsErI	nEs0sErI		76	B			AJ0	AJ0		1.11	0.13875		
643	foiled	fooled	f3l	ful		90	B			AJ0	AJ0		0.3	0.1		
644	free	three	fri	Tri		93	B		AAVE	AJ0	AJ0		0.24	0.08		
645	gastronomical	astronomical	g&str0namIk0l	.&str0namIk0l	coll. astronomical costs	96	B			AJ0	AJ0	C	0.18	0.013846153		
646	gorgeous	gorgers	gOrJ0.s	gOrJ0rz		100	B			AJ0	NN2		0.11	0.015714286		
647	insulted	consulted	.Ins^l?Id	k0ns^l?Id	coll. consult a doctor	120	B			AJ0	AJ0	I	0.19	0.02111111		
648	irritated	irrigated	IrIt4?Id	IrIg4?Id		122	B			AJ0	AJ0		0.66	0.0825		
649	lesbian	Lebanese	lEzbi0n..	lE.b.0niz		136	B			AJ0	AJ0		0.88	0.09777778		
650	oval	awful	5v0l	Of0l		167	B			AJ0	AJ0		0.43	0.1075		
651	quasi	crazy	kw4zI	kr4zI	idiom crazy in the head	188	B			AJ0	AJ0	I+	0.42	0.084		
652	queasy	easy	kwizI	..izI		188	B			AJ0	AJ0		0.21	0.042		
653	queer	dear	kw6r	d.6r	idiom dear to my heart	188	B			AJ0	AJ0	I+	0.45	0.1125		
654	rheumatic	romantic	rum&.?Ik	r5m&n?Ik	idiom incurable romantic	196	B			AJ0	AJ0	C+	0.37	0.04625		
655	torrid	horrid	tarId	harId	quote was a little girl, who had a little...	234	B			AJ0	AJ0	Q+	0.44	0.088		
656	unshed	unsaid	^nSEd	^nsEd	idiom some things are better left unsaid	209	B			AJ0	AJ0	E+	0.09	0.018		
657	clothed	closed	kl5Dd	kl5zd		50	B			AJ0	AJ0		0.44	0.088		
658	mint	meant	mInt	mEnt	idiom I meant well	150	B			VVD	VVD	I+	0.2	0.05		
659	must	messed	m^st	mEst		154	B			VM0	VVD		0.26	0.065		
660	porn	born	pOrn	bOrn	quote a star is born	180	B			NN1	VVN	Q+	0.16	0.04		
661	raffled	ruffled	r&f0ld	r^f0ld	idiom have one's feathers ruffled	190	B			VVN	VVN	E+	0.27	0.045		
662	reaped	raped	ript	r4pt		192	B			VVN	VVN		0.27	0.0675		
663	renegaded	relegated	rEn0g4dId	rEl0g4?Id	idiom relegated to the back seat	195	B			VVN	VVN	I+	0.35	0.03888889		

664	shed	shut	SEd	S^t	idiom some things are better left unsaid	209	B			VVN	VVN	E+	0.38	0.12666667	
665	singed	sinned	sInJd	sIn.d	quote forgive me, father, for I have sinned	211	B			VVN	VVN	Q+	0.34	0.068	
666	forking	fucking	fOrkIN	f^.kIN	coll. fucking around	91	B			VVG	VVG	I	0.31	0.051666666	
667	keel	kill	ki1	kIl		129	B			VVG	VVG		0.1	0.033333335	
668	accost	cost	0kast	.kast		6	B			VVI	VVI		0.04	0.008	+1
669	abomination	denomination	.0bamIn4S0n	dInamIn4S0n	idiom belong to a denomination	6	B			NN1	NN1	I+	0.61	0.055454545	
670	aches	eggs	4ks	Egz	coll. ham and eggs	7	B			NN2	NN2	B	0.44	0.14666666	
671	alder	elder	Old0r	Eld0r	idiom ceremony in front of an elder	10	B			NN1	NN1	E+	0.3	0.06	
672	alligator	elevator	&lIg4?0r	El0v4?0r		10	B			NN1	NN1		0.86	0.1075	
673	anals	annals	4n0lz	&n0lz		84	B	see fart		NN2	NN2		0.27	0.054	
674	angel	angle	4nJ0l	&Ng0l	coll. bisecting angles	30	B	see bisexing		NN1	NN1	C	0.85	0.17	
675	apse	lapse	.&ps	l&ps	coll. lapse of memory	14	B			NN1	NN1	I	0.08	0.02	
676	asses	masses	.&sIz	m&sIz	quote religion is the opiate of the masses	17	B			NN2	NN2	Q+	0.19	0.038	
677	bee	knee	bi	ni	title climb up on my knee, sonny boy	27	B			NN1	NN1	Q+	0.52	0.26	
678	blekfast	breakfast	blEkf0st	brEkf0st		75	B	see election	Chinese	NN1	NN1		0.11	0.01375	
679	brain	plane	br4n	pl4n	quote the rain in Spain ...	216	B	see spay		NN1	NN1	Q+	0.27	0.0675	
680	bungle	jungle	b^Ng0l	J^Ng0l	idiom law of the jungle	38	B			NN1	NN1	E+	0.66	0.11	
681	bus	bush	b^s	bUS	idiom a bird in the hand...	39	B			NN1	NN1	S+	0.45	0.15	
682	buy	rye	b1	rI	title catcher in the rye	39	B			NN1	NN1	Q+	0.66	0.33	
683	cashew	cash who	k&S.u	k&Shu		43	T			NN1	NN1 DTQ		0.07	0.014	
684	caws	calls	kO.z	kOlz	coll. long-distance call	44	B			NN2	NN2	R	0.08	0.02	
685	cells	selves	sEl.z	sElvz	quote the fault, dear Brutus, is not ...	45	B			NN2	NN1	Q+	0.17	0.034	
686	char	chair	Car	C7r	idiom legs over the arm of your chair	46	B			NN1	NN1	E+	0.29	0.096666664	
687	chassis	chaos	C&si.	k4.as	quote the whole world is in a state of chaos	46	B	what?		NN1	NN2	Q+	1.12	0.224	
688	chatter	matter	C&?0r	m&?0r	idiom mind over matter	46	B			NN1	NN1	E+	1	0.2	
689	clings	kings	klINz	k.INz	idiom sport of kings	49	B			NN2	NN2	I+	0.06	0.012	
690	coat	goat	k5t	g5t	idiom sheep are separated from the goats	50	B			NN1	NN1	E+	0.11	0.036666665	
691	comma	coma	kam0	k5m0		52	B	what?		NN1	NN1		0.25	0.0625	
692	convention	invention	k0nvEnS0n	.InvEnS0n	proverb necessity is the mother of invention	54	B			NN1	NN1	S+	0.1	0.011111111	
693	coppice	copies	kapIs	kapIz		55	B			NN1	NN2		0.07	0.014	
694	cores	cause	kOrz	kO.z	coll. cause and effect	55	B			NN2	NN1	B	0.03	0.0075	
695	coroner	corner	kar0n0r	kOr.n0r	proverb death is just around the corner	56	B			NN1	NN1	S+	0.28	0.04	
696	crime	time	kr1m	t.1m	proverb a stitch in time saves nine	214	B	see snitch		NN1	NN1	S+	0.17	0.0425	
697	cruller	color	kr^l0r	k.Ol0r	idiom horse of a different color	59	B			NN1	NN1	E+	0.52	0.074285716	
698	cuneiform	uniform	kjuni0fOrm	.jun.0fOrm	idiom stripped of one's uniform	60	B			NN1	NN1	E+	0.23	0.023	
699	cycle	circle	s1.k0l	s9rk0l	idiom vicious circle	61	B			NN1	NN1	C+	0.33	0.055	
700	defect	effect	dIfEkt	.IfEkt	coll. cause and effect	55	B	see core		NN1	NN1	B	0.11	0.018333333	
701	detention	invention	di.tEnS0n	.InvEnS0n	idiom mother of invention	66	B			NN1	NN1	E+	0.97	0.107777774	
702	dower	hour	d20r	.20r	coll. by the hour	69	B			NN1	NN1	I	0.11	0.0275	
703	dung	tongue	d^N	t^N	idiom speak with forked tongue	72	B			NN1	NN1	E+	0.09	0.03	
704	eel	wheel	.il	wil	idiom put one's shoulder to the wheel	75	B			NN1	NN1	E+	0.13	0.043333333	
705	elk	ilk	Elk	Ilk		76	B			NN1	NN1		0.21	0.07	
706	elm	helm	.Elm	hElm	coll. at the helm	76	B	double pun		NN1	NN1	C	0.07	0.0175	
707	elocution	electrocution	El0k....juS0n	ilEktr0kjuS0n	_def.	76	B			NN1	NN1		0.88	0.06769231	-1
708	emery	memory	.Em0rI	mEm0rI	title Thanks for the Memory	76	B			NN1	NN1	Q+	0.19	0.031666666	
709	emotion	motion	im5S0n	.m5S0n		77	B			NN1	NN1		0.13	0.021666666	+1
710	enamor	a hammer	En&m0r	0h&m0r		77	B	what?		NN1	NN1		0.78	0.13	
711	encyclical	bicycle	.EnsIklIk0l	b1.sIk...0l	quote from Daisy Bell	77	B			NN1	NN1	Q+	0.85	0.07727273	+1
712	error	era	Er0r	ir0.	idiom dawn of an era	79	B			NN1	NN1	E+	0.32	0.08	
713	excess	success	.EksEs	s0ksEs	idiom nothing succeeds like success	81	B			NN1	NN2	E+	0.36	0.06	
714	fade	fate	f4d	f4t	idiom a fate worse than death	83	B			NN1	NN1	E+	0.09	0.03	
715	farce	force	fars	fOrs	coll. tour de force	84	B			NN1	NN1	F	0.24	0.06	
716	fender	feather	fEnd0r	fE.D0r	idiom knock over with a feather	85	B			NN1	NN1	E+	0.15	0.025	
717	fetter	feather	fE?0r	fED0r	proverb bards of a feather ...	24	B	see bard		NN1	NN1	S+	0.23	0.046	
718	flat	fat	fl&t	f.&t	idiom live on the fat of the land	89	B			NN1	NN1	E+	0.06	0.015	
719	fodder	further	fa.d0r	f9rD0r	idiom be further ahead	90	B			NN1	AV0	I+	0.38	0.06333333	
720	fury	jury	fjUrI	J.UrI	idiom trial by jury	94	B			NN1	NN1	I+	1.1	0.22	
721	gal	girl	g&.l	g9rl		95	B	abbreviation gal-gallon		NN1	NN1		0.33	0.0825	
722	gem	jam	JEm	J&m		96	B			NN1	NN1		0.2	0.06666667	
723	gentiles	genitals	JEn.?1lz	JEnIt0lz	quote will of a man ... genitals (Schopenh.)	96	B			NN2	NN2	Q+	0.61	0.07625	-1
724	gibbon	ribbon	gIb0n	rIb0n	quote he promised to buy ... blue ribbons	97	B			NN1	NN1	Q+	0.52	0.104	
725	glue	blue	glu	blu	title Rhapsody in Blue	98	B			NN1	NN1	Q+	0.31	0.10333333	

726	grains	gains	gr4nz	g.4nz	coll. ill-gotten gains	100	B		NN2	NN2	C	0.03	0.006		
727	grass	class	gr&s	kl&s	coll. cook with class	101	B		NN1	NN1	I	0.23	0.0575		
728	greed	creed	grid	krid	idiom regardless of race, color, and creed	101	B		NN1	NN1	Q+	0.1	0.025		
729	guild	guilt	gIld	gIlt	coll. guilt by association	103	B		NN1	NN1	I	0.09	0.0225		
730	hair	fair	h7r	f7r	title Oh Dear what can the matter be...	106	B		NN1	NN1	Q+	0.38	0.12666667		
731	helm	hell	hElm	hEl.	idiom to hell with you	110	B		NN1	NN1	I+	0.14	0.035		
732	hose	house	h5z	h2s	proverb people who live in glass houses...	113	B		NN1	NN1	S+	0.36	0.12		
733	hose	nose	h5z	n5z	idiom win by a nose	113	B		NN1	NN1	E+	0.52	0.17333333		
734	humility	humidity	hjumIl0tI	hjumId0tI	idiom not the heat, but the humidity	114	B		NN1	NN1	E+	0.34	0.037777778		
735	humus	humor	hjum0s	hjum0r	idiom lowest form of humor	114	B		NN1	NN1	I+	0.44	0.07333333		
736	hunk	hank	h^Nk	h&Nk	coll. hank of game	114	B		NN1	NN1	I	0.28	0.07		
737	information	inflammation	Inf.0rm4S0n	InfI0.m4S0n	coll. inflammation in the appendix	119	B		NN1	NN1	I	0.11	0.01		
738	jest	chest	JEst	CEst	idiom play it close to one's chest	124	B		NN1	NN1	E+	0.18	0.045		
739	jinx	links	JINks	lINks		125	B	what?	NN1	NN2		1	0.2		
740	knot	night	nat	n1t	idiom the morning after the night before	131	B		NN1	NN1	E+	0.23	0.07666667		
741	labradors	lavatories	l&br0dOr.z	l&v.0tOrIz		133	B		NN2	NN2		0.41	0.041		-1
742	leeches	breeches	.liCIz	briCIz	idiom caught with one's breeches down	136	B		NN2	NN2	S+	0.32	0.053333335		
743	linen	living	lIn0n	lIvIN	proverb it takes a lot of living ...	138	B		NN1	NN1	S+	1.14	0.228		
744	loom	womb	lum	wum	idiom fruit of the womb	142	B		NN1	NN1	E+	0.44	0.14666666		
745	lore	law	lOr	lO.	idiom possession is nine points of the law	142	B		NN1	NN1	E+	0.03	0.01		
746	mane	name	m4n	n4m	idiom travel under an assumed name	146	B		NN1	NN1	I+	0.25	0.083333336		
747	manse	man's	m&ns	m&nz		146	B	abbr.	NN1	NN1		0.07	0.0175		
748	meter	matter	mi?0r	m&?0r	idiom mind over matter	149	B		NN1	NN1	E+	0.3	0.06		
749	minaret	minute	mIn0rEt	mIn..It	idiom I'll be there in a minute	150	B		NN1	NN1	I+	0.28	0.04		+1
750	molecule	monocle	malIkjul	man0k.0l		151	B		NN1	NN1		0.68	0.085		
751	Monet	money	m5n4	m^nI	idiom marry someone for their money	152	B		NN1	NN1	I+	0.56	0.14	l	
752	money	mummy	m^nI	m^mI	coll. mummy and daddy	152	B		NN1	NN1	B	0.14	0.035		
753	monogamy	monotony	m5nag0mI	m0na?0nI	coll. break the monogamy	152	B		NN1	NN1	I	1	0.125		
754	mulchness	mulch nest	m^lCnIs.	m^lCnEst	coll. mulch nest	154	T		NN1	NN0 NN1	R	0.27	0.03375		
755	nab	gab	n&b	g&b	idiom gift of the gab	156	B		NN1	NN1	E+	1	0.33333334		
756	nag	bag	n&g	b&g	idiom it's in the bag	156	B		NN1	NN1	E+	0.66	0.22		
757	navel	naval	n4v0l	n4v0l	coll. naval engagenment	157	B		NN1	NN1	C	0	0		
758	nonentity	identity	nanEn?ItI	.1dEn?ItI	idiom a case of mistaken identity	160	B		NN1	NN1	E+	0.71	0.078888886		
759	odor	a door	5d0r	0dOr	coll. out of order	164	B		NN1	AT0 NN1	I	0.48	0.12	e	
760	peephole	people	piph5l	pip.0l	idiom power to the people	173	B		NN1	NN1	I+	0.31	0.051666666		
761	persistance	resistance	p0rsIst0ns	rI.zIst0ns	idiom path of least resistance	174	B		NN1	NN1	E+	0.62	0.062		
762	pianist	penis	pi0nIs	pi.nIs	coll. 12-inch penis	175	B		NN1	NN1	R	0.04	0.006666667		+1
763	pickerel	pickle	pIk0r0l	pIk..0l	idiom be in a pickle	175	B		NN1	NN1	I+	0.07	0.01		+1
764	pillow	pillar	pI15.	pII0r	coll. pillar of fire	176	B		NN1	NN1	I	0.27	0.054		
765	pitchers	pictures	pI.C0rz	pIkC0rz	idiom you ought to be in pictures	177	B		NN2	NN2	E+	0.18	0.025714286		
766	pity	city	pI?I	sI?I	coll. the city of New York	177	B		NN1	NN1	I	1	0.25		
767	plaice	face	pl4s	f.4s	title I've grown accustomed to her face	177	B		NN1	NN1	Q+	0.27	0.0675		
768	platitude	altitude	pl&.?Itud	..&l?Itud	coll. altitudes of achievement	178	B		NN1	NN1	I	0.3	0.033333335		
769	plot	pot	plat	p.at	idiom have a chicken in every pot	178	B		NN1	NN1	E+	0.06	0.015		
770	poodle	paddle	pud0l	p&d0l	idiom be up the creek without a paddle	179	B		NN1	NN1	S+	0.3	0.06		
771	poultry	poetry	p5.ltrI	p5I.trI		181	B		NN1	NN1		0.17	0.024285715		
772	puce	pews	pjus	pjuz		184	B		NN1	NN2		0.07	0.0175		
773	reasons	seasons	riz0nz	siz0nz	quote a man for all seasons	193	B		NN2	NN2	Q+	0.66	0.11		
774	rhesus	reason	ris0s	riz0n	quote people's faith in reason	196	B		NN1	NN1	Q+	1.07	0.214		
775	rigids	riches	rIJIdz	rICI.z	idiom from rags to riches	197	B		NN2	NN1	E+	0.29	0.048333332		
776	roam	room	r5m	rum	idiom need lots of room	197	B		NN1	NN1	I+	0.27	0.09		
777	runt	rent	r^nt	rEnt	coll. for rent	200	B		NN1	NN1	I	0.26	0.065		
778	scandal	sandal	sk&nd0l	s.&nd0l		203	B		NN1	NN1		0.1	0.014285714		
779	scarlets	starlets	skarlIts	starlIts	coll. film starlet	119	B	see indigo	NN2	NN2	R	0.14	0.0175		
780	screech	speech	skriC	sp.iC	idiom freedom of speech	204	B		NN1	NN1	I+	0.26	0.052		
781	sentiments	sediments	sEn?0m0nts	sE.d0m0nts	coll. fine sediment	207	B		NN2	NN2	C	0.28	0.028		
782	sequels	equals	sikw0lz	.ikw0lz	idiom treat as equals	207	B		NN2	NN2	I+	0.08	0.011428571		
783	shack	shock	S&k	Sak	idiom to come as a shock	208	B		NN1	NN1	I+	0.23	0.07666667		
784	show van ism	chauvinism	S5v&nIz0m	S5vInIz0m		210	B		NN1	NN1		0.26	0.028888889		
785	shrivelry	chivalry	SrIv0lrI	S.Iv0lrI	coll. act of chivalry	210	B		NN1	NN1	I	0.03	0.00375		
786	slug	rug	sl^g	.r^g	idom sweep something under the rug	213	B		NN1	NN1	E+	0.19	0.0475		
787	smacks	snacks	sm&ks	sn&ks	idiom drop in for snacks	213	B		NN2	NN2	I+	0.11	0.022		

788	spaying	Spain	sp4IN	sp4.n	quote the rain in Spain ...	216	B			NN1	NP0	Q+	0.29	0.058		+1
789	spree	tree	spri	.tri	quote a partridge in a pear tree	218	B			NN1	NN1	Q+	0.26	0.065		
790	stairs	stars	st7rz	starz	idiom avoid the stars	218	B	see also cell		NN1	NN2	I+	0.29	0.058		
791	stilling	killing	stIlIN	.kIlIN	idiom kill two birds with one stone	53	B	see cone		NN1	NN1	E+	0.24	0.04		
792	stork	storm	stOrk	stOrm	idiom calm before the storm	221	B			NN1	NN1	I+	0.66	0.132		
793	Strauss	stress	str2s	strEs	coll. stress and strain	221	B			NN1	NN1	B	0.3	0.06		
794	sty	sky	st1	sk1	quote a goldmine in the sky	222	B			NN1	NN1	Q+	0.16	0.053333335		
795	Swede	sweet	swid	swit	idiom take the bitter with the sweet	224	B			NN1	NN1	E+	0.09	0.0225		
796	her sylph	herself	sIlf	sElf	idiom keep something to oneself	224	B			DPS NN1	PNX	I+	0.2	0.05		
797	Talmud	tall mud	talmUd	tOlm^d		226	B			NN1	AJ0 NN1		0.53	0.08833333		
798	thaw	draw	T.O	drO	idiom quick on the draw	229	B			NN1	NN1	I+	0.7	0.23333333		
799	tile	tale	t1l	t4l	title the tales from the Vienna woods	232	B			NN1	NN1	Q+	0.27	0.09		
800	tine	time	t1n	t1m	idiom fall on hard times	232	B			NN1	NN1	E+	0.14	0.046666667		
801	trail	trial	tr4.l	tr10l	idiom learn by trial and error	234	B			NN1	NN1	I+	0.33	0.066		
802	trend	tent	trEnd	t.Ent	idiom right around the tent	235	B	WS had trend-end		NN1	NN1	E+	0.12	0.024		
803	vampire	empire	v&mp10r	.Emp10r	idiom the sun never sets on the Br. Empire	242	B			NN1	NN1	Q+	0.53	0.07571428		
804	vault	fault	vOlt	fOlt	idiom generous to a fault	243	B			NN1	NN1	E+	0.17	0.0425		
805	vesper	whisper	vEsp0r	wIsp0r	idiom talk above a whisper	244	T			NN1	NN1	I+	0.29	0.048333332		
806	VIP	way I pee	vi1pi	w41pi		245	B	rel. clause		NN1	NN1 PNP VVB		0.35	0.07		
807	vituperation	recuperation	vItup0r4S0n	rIkup0r4S0n	coll. power of recuperation	246	B			NN1	NN1	I	0.6	0.054545455		
808	wane	plain	.w4n	pl4n	quote the rain in Spain ...	249	B			NN1	NN1	Q+	0.49	0.1225		
809	wedlock	hemlock	wEdlak	hEmlak		250	B			NN1	NN1		0.82	0.13666667		
810	weevil	evil	wiv0l	.iv0l	idiom the lesser of two evils	251	B			NN1	NN1	E+	0.11	0.022		
811	whisker	whisper	wIsk0r	wIsp0r	idiom talk below a whisper	252	B			NN1	NN1	I+	0.23	0.038333334		
812	whorescopes	horoscopes	hOr.sk5ps	har0sk5ps		253	B	P: non-word		NN2	NN2		0.32	0.035555556		-1
813	wife	wise	w1f	w1z	idiom a word to the wise	253	B			NN1	NN1	E+	1	0.33333334		
814	windrow	window	wIndr5	wInd.5	quote serenade ... 'neath a fair lady's window	254	B			NN1	NN1	Q+	0.03	0.005		
815	yearning	learning	j9rnIN	l9rnIN	coll. higher learning	259	B			NN1	NN1	C	0.25	0.041666668		
816	bourbon	Burden	burb0n	burd0n	idiom beast of burden	34	B	overlooked by WS		NN1	NN1	I+	0.21	0.035		
817	cervix	service	s9rvIks	s9rvI.s	idiom at your service	67	B	see p. 38		NN1	NN1	I+	0	0		
818	platitude	latitude	pl&d0tjud	.l&d0tjud	coll. longitude and latitude	178	B			NN1	NN1	B	0	0		
819	Acapulco	a capella	ak0pUlk5	ak0pEl.0	_def.	6	ne		Italian	NP0	PRP NN1		0.63	0.07875		
820	Aeneid	he need	.iniId	hini.d		8	T			NP0	PNP VVB		0.17	0.028333334		
821	agate	I got	&gIt	1gat		9	T	Agate=name of font?		NP0	PNP VMB		0.5	0.125	l	
822	Andalusian	and illusion	&nd0luZ0n	0ndIluZ0n		13	T	what?		NP0	CJC NN1		0.23	0.025555555		
823	Argus	our guess	arg0s	argEs	idiom our guess is as good as yours	15	B	atrocious		NP0	AT0 NN1	E+	0.23	0.046		
824	armageddon	I'm a-gettin'	arm0gEd0n	1.m0gE?In	idiom I'm getting out of here	16	B			NP0	PNP VBB VVG	I+	0.38	0.04222222		
825	Atlas	at last	&tl0s	0tl&s	idiom at last, we are alone	18	ne			NP0	PRP AV0	Q+	0.47	0.094	e	
826	Avon	havin'	.1v0n	h&vIn	idiom are you having any fun	20	T			NP0	VHG	S+	0.36	0.072		
827	Baal	bail	b40l	b4.l		21	ne			NP0	VVB		0.04	0.01		
828	Bangladesh	bang the dish	baNgl0dES	b&ND.0dIS		23	T			NP0	VVI AT0 NN1		0.89	0.09888889		
829	Caruso	Crusoe	k0rus5	k.rus5	_def.	43	B			NP0	NP0		0.04	0.006666667		
830	Champlain	champaigne	S&mpl4n	S&mp.4n	_def.	45	ne		Chinese	NP0	NN1		0.06	0.008571428	e	
831	Champlain	chomping	S&mpl4n	Camp.IN	idiom horses chomping at the bit	45	T			NP0	VVG	Q+	0.86	0.122857146		
832	Chianti	can't ye	kiant.i	k.&ntj4		47	B	WS:Chianti-can't you, see rum		NP0	VMB XX0 PNP		0.68	0.09714286		
833	Dresden	dressed in	drEzd0n	drEstIn		70	T			NP0	VVN AVP		0.16	0.022857143		
834	dunce	dance	d^ns	d&ns	idiom may I dance with you	72	T			NP0	VVI	I+	0.28	0.07		
835	Essex	ass aches	Es0ks	&s4ks		79	ne			NP0	NN1 VVZ		0.48	0.096		
836	Hallelujah	Honolulu	h&l0luj0	han0lulu	_def.	106	ne			NP0	NP0		1.07	0.13375		
837	Harvard	harder	harv0rd	hard0r.	coll. try harder	108	T			NP0	AV0	I	0.55	0.07857143		
838	Hogmanay	hog many	hagm0n4	hagmEnI		112	T			NP0	VVI PN1		0.51	0.07285714	l	
839	Kodak	Koran	k5d&k	k0ran	_def.	132	ne			NP0	NP0		1.41	0.282	e	
840	koodoo	could you	kud.u	kUdju	idiom brother, could you spare a dime	132	T			NP0	VBB PNP	Q+	0.38	0.076		
841	Kumquat	come quick	k^mkwat	k^mkwIk		132	T			NP0	VVB AV0		0.44	0.06285714		
842	Manhasset	man has it	m&nh&sIt	m&nh&zIt		157	T	see Nantucket		NP0	NN1 VHZ PNP		0.07	0.00875		
843	Mecca	maker	mEk0.	m4k0r	idiom meet one's maker	148	T			NP0	NN1	E+	0.32	0.064		
844	Nantucket	Nan took it	n&nt^kIt	n&ntUkIt		157	T			NP0	NP0 VVD PNP		0.28	0.035		
845	Nero	near	n6r5	n6r.	idiom the end is near	158	T			NP0	AJ0	I+	0.24	0.06		+1
846	nurture	nature	n9rC0r	n4.C0r	coll. mother nature	162	B			NP0	NN1	R	0.33	0.055		
847	Parthenon	passing on	parT0nan	p&.sInan	coll. an idea worth passing on	170	T			NP0	VVG AVP	P	0.32	0.04		
848	Pontius	conscious	p^nS0s	kanS0s	coll. semi-conscious	179	T			NP0	AJ0	R	0.5	0.083333336		
849	Seoul	sold	s5l.	s5ld		207	T			NP0	VVN		0.09	0.0225		

850	Thinatra	Sinatra	TInatr0	sInatr0		230	ne		lisp	NP0	NP0		0.11	0.015714286		
851	Tibet	to bet	tIbEt	t0bEt	title to be your life	231	T			NP0	TO0 VVI	Q+	0.01	0.002		
852	Uruguay	your way	jUr0gw4	jOr..w4	proverb you go your way and I'll go mine	241	T			NP0	DPS NN1	S+	0.51	0.07285714		+1
853	Pawtucket	Pa took it	pOtakIt	patUkIt		157	T	see Nantucket		NP0	NN1 VVD PNP		0	0		
854	Faust	first	f2.st	f9rst	quote first in war, first in peace, ...	85	B			NP0	AV0	Q+	0.34	0.068		
855	Harvard	our word	harv0rd	.arw0rd	idiom our word against yours	108	B			NP0	DPS NN1	I+	0.16	0.022857143		
856	Joyce	choice	J3s	C3s	coll. unfortunate choice	126	B			NP0	NN1	C	0.18	0.06		
857	Kahluah	clue	k0l8	k.lu	idiom have a clue	128	B			NP0	NN1	I+	0.33	0.0825		+1
858	Iraq	wreck	Irak	.rEk	idiom things are going from wreck to ruin	122	B			NP0	NN1	E+	0.37	0.0925		+1
859	Genoan	genuine	JEn0w0n	JEnu.In	coll. the genuine article	96	B			NP0	AJ0	C	0.39	0.055714287		
860	Korea	carrier	k0ri0.	k&ri0r	coll. carrier girl	132	B			NP0	NN1	R	0.28	0.046666667	l	
861	Orly	early	OrlI	9rlI	proverb the early bird gets the worm	250	B	see warm		NP0	AJ0	S+	0.29	0.0725		
862	Ruth	truth	.rUT	trUT	proverb truth is stranger than fiction	200	B			NP0	NN1	S+	0.07	0.0175		
863	acclaim	a clam	0kl4m	0kl&m	coll. clam bake	6	T			NN1	AT0 NN1	R	0.27	0.054		
864	adder	utter	&d0r	^?0r	coll. utter confusion	8	B			NN1	AJ0	C	0.39	0.0975		
865	addict	attic	&dIkt	&?Ik.	coll. attic rendevouz	8	B			NN1	NN1	R	0.19	0.038		
866	almond	arm and	a.m0nd	arm0nd	idiom cost an arm and a leg	11	B	no 2nd script		NN1	NN1 CJC	E+	0.04	0.006666667		
867	ambivalence	ambulance	&mbIv0l0ns	&mb.j0l0ns	coll. ambulance chasers	12	B	syllable lost		NN1	NN1	R	1.1	0.11		+1
868	antic	antique	&n?Ik	&ntik	coll. antique collector	13	B			NN1	NN1	R	0.36	0.072		
869	author	other	OT0r	^D0r		19	B	sem. field		NN1	AJ0		0.51	0.1275		
870	bazaar	bizzare	b0zar	bizar	coll. bizarre experience	25	B			NN1	AJ0	C	0.26	0.052		
871	beaver	fever	biv0r	fiv0r	coll. feaver pitch	26	B			NN1	NN1	R	0.66	0.132		
872	belly	very	bElI	vErI		28	B	syntax prevent much else		NN1	AV0		0.29	0.0725		
873	blight	white	bl1t	.w1t	coll. white supremacy	31	B			NN1	AJ0	R	0.65	0.1625		
874	bottle	battle	ba?0l	b&?0l	coll. battle fatigue	33	B	WS confused w/ metal fatigue		NN1	NN1	R	0.18	0.036		
875	bourbon	urban	burb0n	.9rb0n	coll. urban renewal	34	B	see rum		NN1	NN1	C	0.51	0.085		
876	cedar	see the	sid0r	siD0.	coll. cedars of Lebanon	44	B			NN1	VVB AT0	I	0.08	0.016		
877	cervix	service	s9rvIks	s9rvI.s	coll. service station	45	B			NN1	NN1	R	0.1	0.014285714		
878	clue	glue	klu	glu		50	B			NN1	NN1		0.11	0.036666665		
879	cod	cold	ka.d	k5ld	coll. cold war	51	B			NN1	AJ0	R	0.33	0.0825		
880	creche	crash	kr4S	kr&S	coll. crash diet	58	B		French	NN1	NN1	R	0.27	0.0675		
881	crossroad	crossword	krasr5.d	krasw9rd	coll. crossword puzzle	59	B			NN1	NN1	R	0.55	0.06875		
882	drogue	drug	dr5g	dr^g	coll. drug store	70	B			NN1	NN1	R	0.28	0.07		
883	due	blue	d.u	blu	coll. blue horizon	71	B			NN1	AJ0	C	0.46	0.15333334		
884	earring	hearing	.6rIN	h6rIN	coll. hearing aid	74	B			NN1	AJ0	C	0.07	0.014		
885	elf	shelf	.Elf	SElf	coll. shelf space	76	B			NN1	NN1	R	0.52	0.13		
886	epoch	pock	ipak	.pak	coll. pock-marked	78	B			NN1	NN1	R	0.13	0.0325	e	+1
887	eunuch	unique	jun0k	junik	coll. unique experience	80	B			NN1	AJ0	C	0.26	0.052	e	
888	eyrie	eerie	Eri	iri	coll. eerie feeling	82	B			NN1	AJ0	C	0.29	0.096666664		
889	fat	cat	f&t	k&t	quote the fog comes on ...	68	B	quote identified correctly?		NN1	NN1	Q+	0.44	0.14666666		
890	feed	speed	.fid	spid	coll. speed limit	85	B			NN1	NN1	R	0.33	0.0825		
891	fig	big	fIg	bIg	idiom I'm a big boy now	87	B			NN1	NN1	E+	0.38	0.12666667		
892	file	vile	f1l	v1l	coll. vile idea	87	B			NN1	AJ0	C	0.14	0.046666667		
893	flea	pea	fli	p.i	idiom thick as pea soup	89	B			NN1	NN1	E+	0.29	0.096666664		
894	floor play	foreplay	flOr	f.Or	coll. foreplay	89	B	just morpheme		NN0 NN1	NN1	R	0.06	0.015		
895	fun	fund	f^n.	f^nd	coll. fund raiser	94	B			NN1	NN1	R	0.09	0.0225		
896	genital	gentle	JEnIt0l	JEn.?0l	coll. gentle persuasion	96	B			AV0	AV0	C	0.27	0.03857143		+1
897	glass	grass	gl&s	gr&s	coll. grass widow	98	B			NN1	NN1	R	0.11	0.0275		
898	graft	graf (Zeppelin)	graft	graf.	coll. Graf Zeppelin	100	B	title ~ PropR		NN1	NP0	F	0.07	0.014		
899	gust	ghost	g^st	g5st	coll. ghost story	103	B			NN1	NN1	R	0.29	0.0725		
900	heir	hair	.7r	h7r	coll. heir removal	110	B			NN1	NN1	R	0.07	0.023333333		
901	herd	hard	h9rd	hard	idiom do it the hard way	110	B			NN1	AJ0	E+	0.29	0.0725		
902	hippie	happy	hIpI	h&pI	title happy birthday to you	111	B			NN1	AJ0	Q+	0.28	0.07		
903	homily	family	hamlII	f&mlII	coll. family woman	112	B			NN1	NN1	R	0.56	0.093333334		
904	house	horse	h2.s	hOrs	coll. horse opera	114	B			NN1	NN1	R	0.33	0.0825		
905	hurt	heart	h9rt	hart	coll. heart-to-heart	114	B			NN1	NN1	B	0.29	0.0725		
906	interval	integral	In?0rv0l	In?0gr0l	coll. an integral part	121	B			NN1	AJ0	C	1.44	0.18		
907	ire	higher	.10r	h10r	coll. higher things	122	B			NN1	AJ0	C	0.07	0.0175		
908	jam	damn	J&m	d&m	coll. one thing after another	124	B			NN1	AJ0	I	0.66	0.22		
909	jenny	any	JEnI	.EnI	coll. spinning jenny	124	B			NN1	AT0	C	0.34	0.085		
910	jowl	jaw	J2l	JO.	coll. cheek by jowl	125	B			NN1	NN1	B	0.35	0.11666667		
911	jungle	jingle	J^Ng0l	JINg0l	title jingel bells	127	B			NN1	NN1	Q+	0.29	0.048333332		

912	jut	jet	J^t	JEt	coll. jet set	127	B			NN1	NN1	R	0.26	0.086666666		
913	lay	day	l4	d4	coll. day preacher	134	B			NN1	NN1	R	0.34	0.17		
914	lemming	lemon	lEmIN	lEm0n	coll. lemon juice	136	B			NN1	NN1	R	0.2	0.04		
915	lentil	Lent till	lEn?0l	lEntIl		136	B			NN1	NP0 PRP		0.21	0.035		
916	lingo	bingo	lINg5	bINg5	coll. bingo game	139	B			NN1	NN1	R	0.66	0.132		
917	litter	little	lI?0r	lI?0l	idiom every little bit helps	140	B			NN1	AJ0	I+	0.13	0.026		
918	litter	bitter	lI?0r	bI?0r	idiom bitter pill	140	B	see also Swede		NN1	AJ0	I+	0.66	0.132		
919	locust	low cost	l5k0st	l5kast	coll. low-cost food	141	B			NN1	AJ0 NN1	C	0.26	0.043333333	e	
920	lore	lower	lO.r	l50r	coll. lower body	142	B			NN1	AJC	C	0.35	0.0875		
921	lust	just	l^st	J^st		143	B			NN1	AJ0		1	0.25		
922	mink	mint	mINk	mInt	coll. mint condition	150	B			NN1	AJ0	C	0.33	0.0825		
923	nacelle	nasal	n0sEl	n4z0l	coll. nasal drops	156	B			NN1	AJ0	C	0.61	0.122	l	
924	nay	May	n4	m4	coll. May day celebration	158	B			NN1	NP0	I	0.14	0.07		
925	neap	neat	nip	nit	coll. neat trick	158	B			NN1	AJ0	C	0.18	0.06		
926	nit	net	nIt	nEt	coll. net loss	159	B			NN1	NN1	R	0.2	0.06666667		
927	noodle	new dill	nud0l	nudIl		160	B	post-punch explanation		NN1	AJ0 NN1		0	0	e	
928	nook	book	nUk	bUk	coll. bookworm	160	B			NN1	NN1	R	0.66	0.22		
929	notion	potion	n5S0n	p5S0n	coll. love potion	161	B			NN1	NN1	R	0.66	0.132		
930	nudge	fudge	n^J	f^J	coll. fudge sundae	162	B			NN1	NN1	R	1	0.33333334		
931	nymph	lymph	nImf	lImf	coll. lymph glands	162	B			NN1	NN1	R	0.23	0.0575		
932	owl	old	5l.	5ld	proverb there's no fool like an old fool	167	B			NN1	AJ0	S+	0.09	0.03		
933	peasant	pleasant	p.Ez0nt	plEs0nt		172	B	what?		NN1	AJ0		0.15	0.021428572		
934	pheasant	peasant	fEz0nt	pEz0nt	quote rogue and peasant slave (Hamlet)	174	B			NN1	NN1	Q+	0.23	0.038333334		
935	plinth	prince	plInT	prInT	coll. prince charming	178	B			NN1	NN1	I	0.11	0.022		
936	Poe	poor	p5.	p8r	idiom the poor man's something	178	B			NN1	AJ0	I+	0.34	0.11333334		
937	quiff	cliff	kwIf	kIIf	coll. cliff hanger	189	B			NN1	NN1	R	0.25	0.0625		
938	rabbit	rapid	r&bIt	r&pId	coll. rapid transit	190	B			NN1	AJ0	C	0.25	0.05		
939	raccoon	wreck and (ruin)	r&kun	rEk0n	idiom go to wreck and ruin	190	B			NN1	NN1 CJC	E+	0.45	0.09	l	
940	raid	great	.r4d	gr4t	title the Great Skelton's music guide	191	B			NN1	AJ0	Q+	0.47	0.1175		
941	rail	real	r4l	ril	idiom the real thing	191	B		NYC	NN1	AJ0	I+	0.29	0.096666664		
942	rear	near	r6r	n6r	coll. in the near future	193	B			NN1	AJ0	I	0.31	0.10333333		
943	remorse	Morse	rImOrs	..mOrs	coll. morse code	194	B			NN1	NN1	R	0.13	0.021666666		+1
944	rube	rude	rub	rud	coll. rude shock	199	B			NN1	AJ0	C	0.21	0.07		
945	rye	right	r1.	r1t	coll. dry sense of humor	200	B			NN1	AJ0	I	0.07	0.023333333		
946	sailfish	selfish	s4lfIS	sElfIS	coll. selfish motives	202	B			NN1	AJ0	C	0.26	0.043333333		
947	samovar	some of our	s&m0var	s^m0var		202	B			NN1	PN1 PRP DPS		0.27	0.03857143		
948	scheme	steam	skim	stim	coll. steam engine	204	B			NN1	NN1	R	0.14	0.035		
949	scrotum	totem	skr5?0m	.t.5?0m	coll. totem pole	205	B			NN1	NN1	R	0.25	0.035714287		
950	seek	sick	sik	sIk	coll. sick joke	206	B			NN1	AJ0	C	0.1	0.033333335		
951	shad	sad	S&d	s&d	coll. a sad occasion	208	B			NN1	AJ0	C	0.09	0.03		
952	shark	shock	Sark	Sa.k	coll. shock absorbers	208	B			NN1	NN1	R	0.03	0.0075		
953	shay	jay	S4	J4	coll. jay walking	208	B			NN1	NN1	I	0.38	0.19		
954	shock	shop	Sak	Sap	coll. shop talk	209	B			NN1	NN1	R	0.23	0.07666667		
955	sincere	sear	sIns6r	...s6r	coll. sear sucker suit	211	B	what?		AJ0	NN1	R	0.28	0.046666667		+1
956	snore	shore	snOr	S.Or	coll. shore duty	214	B			NN1	NN1	R	0.27	0.0675		
957	spaniel	Spanish	sp&nj0l	sp&n.IS	coll. Spanish Armada	216	B			NN1	AJ0	R	1.1	0.15714286		
958	spice	space	sp1s	sp4s	coll. aerospace industry	217	ne	just morpheme		NN1	NN1	R	0.27	0.0675		
959	steel	still	stil	stIl	idiom we can all still be friends	219	B			NN1	AV0	E+	0.1	0.025		
960	stork	stark	stOrk	start	coll. stark naked	221	B			NN1	AV0	C	0.39	0.078		
961	suite	seat	swit	s.it	coll. backseat driver	222	B			NN1	NN1	R	0.11	0.0275		
962	swish	shish	swIS	S.IS	coll. shish kebab	224	B			NN1	NN1	F	0.28	0.07		
963	tapir	tapered	t4p0r.	t4p0rd		226	B	quote?		NN1	AJ0		0.09	0.015		
964	thigh	bi	T1	b1	coll. bi-focals	230	B	just morpheme		NN1	UNC	R	1	0.5		
965	three	tree	Tri	tri	quote a partridge in a pear tree	230	B			CRD	NN1	Q+	0.2	0.06666667		
966	tin	thin	tIn	TIn	idiom the thin red line	232	B			NN1	AJ0	E+	0.18	0.06		
967	tongue	kung (fu)	t^N	kUN	coll. kung fu	233	B	just morpheme		NN1	UNC	F	0.44	0.14666666		
968	totter	tartar	ta.?0r	tar?0r	coll. tartar sauce	235	B			NN1	NN1	R	0.04	0.006666667		
969	tour	whore	t8r	hOr	coll. whore house	234	B			NN1	NN1	R	0.73	0.24333334		
970	tram	damn	tr&m	d.&m	idiom one damn thing after another	235	B			NN1	AJ0	I+	0.14	0.035		
971	trip-tease	strip-tease	.trIptiz	strIptiz	coll. striptease	236	B			NN1	NN1	R	0.1	0.0125		
972	troll	droll	tr5l	dr5l	idiom a droll sense of humor	236	B			NN1	AJ0	E+	0.11	0.0275		
973	tweet	sweet	twit	swit	coll. sweet music	238	B			NN1	AJ0	C	0.44	0.11		

974	vamp	lamp	v&mp	l&mp	coll. lamp light	242	B			NN1	NN1	R	0.44	0.11		
975	vest	vast	vEst	v&st	coll. vast mistake	244	B			NN1	AJ0	C	0.2	0.05		
976	vice	nice	v1s	n1s	idiom what's a nice girl like you doing ...	245	B			NN1	AJ0	E+	0.66	0.22		
977	visa	these are	vis0.	Dis0r		246	ne	what?		NN1	DT0 VBB		0.42	0.084		
978	vise	wise	v1s	w1z	coll. wise guy	246	B			NN1	AJ0	C	0.15	0.05		
979	vowel	bowel	v20l	b20l	coll. bowel movement	247	B			NN1	NN1	R	0.44	0.11		
980	whore	horror	hOr..	hOr0r	coll. horror show	253	B			NN1	NN1	R	0.11	0.022		-1
981	yard	hard	jard	hard	idiom it's a hard life	259	B			NN1	AJ0	E+	0.44	0.11		
982	yarn	darn	jarn	darn	coll. any damn thing	259	B			NN1	AJ0	I	1	0.25		
983	bust	bus	bast	bas.	coll. bus stop	39	B			NN1	NN1	R	0	0		
984	pun	pen	pan	pEn	coll. pen pal	184	B			NN1	NN1	R	0	0		
985	bourbon	urban	b9rb0n	.9rb0n	coll. urban life	199	B			NN1	AJ0	C	0	0		
986	sheik	silk	Si.k	sIlk	coll. silk veil	209	B	collocation recoverability		NN1	NN1	R	0	0		
987	Rhodes scholared	rose colored	r5dskal0rd	r5.zk^l0rd	idiom rose-colored glasses	196	B			NP0 NN1	AV0 AJ0	I+	0.39	0.039		
988	cane chew	can't you	k4nCU	k&nC0	title can't you hear me calling, ...	42	T			NN0 NN1	VM0 XX0 PNP	Q+	0.56	0.112		
989	highway cop	I wake up	h1w4kap	.1w4k^p		111	T			NN0 NN1	PNP VVB AVP	P	0.29	0.04142857		
990	liner pier	lyin' up here	l1.n0rp6r	l1In0.p6r		139	T			NNO NN1	VVG P AV0		0.14	0.015555556		
991	H heaver	achiever	4Chiv0r	0C.iv0r	coll. late achiever	105	B			ZZ0 NN1	NN1	C	0.36	0.05142857		
992	opera-tunist	opportunist	ap0r0tunIst	ap0r.tunIst		236	B	see tune only morpheme		NN0 NN1	NN1		0.04	0.0036363637		
993	Oxo-gin	oxygen	aks0JIn	aksIJ0n	coll. breathe oxygen	167	B			NP0 NN1	NN1	I	0.01	0.0014285714		
994	knock need	knock knee	naknid	nakni.	coll. knock knee	158	B	WS		NN0 NN1	NN0 NN1	R	0.11	0.018333333		
995	Nostrum Damus	Nostradamus	nastr0md4m0s	nastr0.d4m0s		161	B			UNC NP0	NP0		0.14	0.011666667		
996	Umber Ella	umbrella	^mb0rEl0	^mb.rEl0	_def.	239	ne			NN0 NP0	NN1		0.04	0.005		+1
997	clip-sew	calipso	k.lIps5	k0lIps5	coll. calipso music	49	B			NN1	NN1	R	0.07	0.01		-1
998	clock	cock	klak	k.ak	coll. cocksucker	49	B	I knew it was out there		NN1	NN1	R	0.06	0.015		
999	clods	gods	kladz	g.adz	idiom chariot of the gods	49	B			NN1	NN1	I+	0.17	0.034		
1000	bugger	beggar	b^g0r	bEg0r	title beggar's opera	37	B			NN1 POS	NN1 POS	Q+	0.26	0.052		
1001	harpies	Harper's	harpI.z	harp0rz	title Harper's Bazaar	107	B			NN1	NP0 POS	Q+	0.05	0.007142857		
1002	sheik	sheep	Sik	Sip	idiom wolf in sheep's clothing	209	B			NN1 POS	NN1 POS	E+	0.23	0.07666667		
1003	teensters'	teamsters'	tinst0rz	timst0rz	coll. teamsters' union	228	B			NN1 POS	NN1 POS	R	0.14	0.0175		
1004	ale	all	4l	Ol	proverb all's well that ends well	10	B	see ale-hell		NN1	NN1	S+	0.29	0.145		
1005	artichoke	architect	ar?IC5k.	arkItEkt		16	B	what?		NN1	NN1		1.4	0.175		
1006	bitch	stitch	.bIC	stIC	proverb a stitch in time saves nine	30	B			NN1	NN1	S+	0.54	0.135		
1007	bloat	boat	bl5t	b.5t	title row, row, row your boat	31	ne	env. adjusted		NN1	NN1	Q+	0.06	0.015		
1008	bride	bread	br1d	brEd	idiom which side one's bread is buttered on	25	B	see barter		NN1	NN1	E+	0.3	0.075		
1009	burp	bird	b9rp	b9rd	idiom a little bird told me	38	B			NN1	NN1	S+	0.44	0.11		
1010	conversation	conservation	kanv0rs4S0n	kans0rv4S0n		54	B	metathesis!		NN1	NN1		1.66	0.1509091		
1011	cranes	rain	kr4nz	.r4n.	title the rain in spain ...	57	B	sg-pl adjustment		NN2	NN1	Q+	0.35	0.07		
1012	cut	cup	k^t	k^p	idiom my cup runneth over	61	B			NN1	NN1	Q+	0.31	0.10333333		
1013	dace	days	d4s	d4z	idiom happy days are here again	62	B			NN1	NN2	Q+	0.07	0.023333333		
1014	denier	then you	d0n.6r	D0nj0.		65	B			NN1	CJS PNP		0.47	0.07833333	l	
1015	dog	fog	dag	fag	quote the fog comes on ...	68	B	where?		NN1	NN1	Q+	1	0.33333334		
1016	era	error	6r0.	Er0r	coll. Churchill Era	78	B			NN1	NN1	R	0.33	0.0825		
1017	fertilization	naturalization	f9rt..Il0z4S0n	n&.C0r0l0z4S0n	_def.	86	B	what?		NN1	NN1		1.48	0.105714284		-1
1018	flame	fame	fl4m	f.4m	idiom s.o.'s fame has spread	88	B	see also indigo		NN1	NN1	I+	0.06	0.015		
1019	flute	fruit	flut	frut		90	B		Chinese	NN1	NN1		0.11	0.0275		
1020	gas	guess	g&s	gEs	idiom your guess is as good as mine	96	B			NN1	NN1	E+	0.19	0.06333333		
1021	groom	broom	grum	brum	proverb a new broom sweeps clean	102	B			NN1	NN1	S+	0.31	0.0775		
1022	gull	lull	g^l	l^l	idiom lull in the conversation	103	B			NN1	NN1	I+	1	0.33333334		
1023	hams	hands	h&m.z	h&ndz	idiom all hands on deck	106	B			NN2	NN1	E+	0.2	0.04		
1024	herd	bird	h9rd	b9rd	proverb a bird in the hand ...	110	B			NN1	NN1	S+	1	0.25		
1025	honey	money	h^nI	m^nI		113	B	what?		NN1	NN1		0.52	0.13		
1026	interloper	antelope	In?0rl5p0r	&n?I.l5p..	idiom where the deer and antelope play	121	B			NN1	NN1	E+	0.38	0.038		+1
1027	lays	days	l4z	d4z	title oh, the days dwindle down ...	134	B	see also May		NN2	NN2	Q+	0.34	0.11333334		
1028	lesion	legion	liZ0n	liJ0n	coll. French Foreign Legion	136	B			NN1	NN1	R	0.31	0.062		
1029	lewd	Lord	lu.d	lOrd	quote the lord giveth, and the lord ..	137	B			NN1	NN1	Q+	0.33	0.0825		
1030	lice	lights	l1.s	l1ts	idioms head lice head lights	137	B			NN1	NN2	R	0.07	0.0175		
1031	loining	learning	l3.nIN	l9rnIN	quote little learning is a dangerous thing (Pope)	141	B			NN1	NN1	Q+	0.33	0.055		
1032	lynch	inch	lInC	.InC	proverb an inch is as good as a mile	143	B			NN1	NN1	S+	0.06	0.015		
1033	malicious	militia	m0lIS0s	mIlIS0.		145	B			NN1	NN1		0.08	0.011428571		
1034	maneuver	manure	m0nuv0r	m0n8..r	coll. horse manure	146	B			NN1	NN1	R	0.67	0.095714286		
1035	mines	minds	m1n.z	m1ndz	proverb great minds think alike	220	B	see stink		NN2	NN1	S+	0.09	0.018		

1036	nail	mail	n4l	m4l	idiom the mail must go through	156	B			NN1	NN1	Q+	0.14	0.046666667	
1037	nudes	news	nudz	nu.z	title no news is good news (Wodehouse)	162	B			NN2	NN2	Q+	0.11	0.0275	
1038	oak	joke	.5k	J5k	idiom the joke's on you	163	B			NN1	NN1	I+	0.52	0.17333333	
1039	pain	rain	p4n	r4n	quote the rain in Spain ...	216	B	see spay		NN1	NN1	Q+	0.52	0.17333333	
1040	parking	parting	parkIN	par?IN	idiom parting is such sweet sorrow	170	B			NN1	NN1	E+	0.66	0.11	
1041	peepers	people	pip0rz	pip0l.	proverb people who live in glass houses...	113	B	see hose		NN2	NN1	S+	0.38	0.06333333	
1042	prose	rose	pr5z	.r5z	proverb a rose by any other name....as sweet	183	B			NN1	NN1	S+	0.16	0.04	
1043	pup	cup	p^p	k^p	quote my cup runneth over	185	B			NN1	NN1	Q+	0.24	0.08	
1044	rube	rub	rub	r^b	quote aye, there's the rub	199	B			NN1	NN1	Q+	0.28	0.093333334	
1045	rum	room	r^m	rUm	quote there's more room to fight	199	B	syntax not identical ?!		NN1	NN1	Q+	0.28	0.093333334	
1046	seasons	seasonings	siz0..nz	siz0nINz		205	B			NN2	NN2		0.35	0.04375	-1
1047	servant	serpent	s9rv0nt	s9rp0nt		207	B			NN1	NN1		0.66	0.09428571	
1048	snail	trail	sn4l	tr4l	title there's a long, long trail a-winding	214	B			NN1	NN1	Q+	0.63	0.1575	
1049	snitch	stitch	snIC	stIC	proverb a stitch in time saves nine	214	B			NN1	NN1	S+	0.34	0.085	
1050	stalk	stork	stO.k	stOrk		218	B			NN1	NN1		0.04	0.008	
1051	strain	rain	str4n	..r4n	quote into every life a little rain must fall	221	B			NN1	NN1	Q+	0.15	0.03	
1052	teal	tail	til	t4l		227	B			NN1	NN1		0.27	0.09	
1053	thing	think	TIN.	TINk	quote I think that I shall never see ...	230	B			NN1	VVB	Q+	0.18	0.045	
1054	tool	fool	tul	ful	proverb there's no fool like an old fool	233	B			NN1	NN1	S+	1	0.33333334	
1055	tufts	tough	t^fts	t^f..	proverb when the going gets tough ...	237	B			NN2	NN1	S+	0.15	0.03	
1056	unction	action	^NkS0n	&.kS0n	proverb actions speak louder than words	240	B			NN1	NN1	S+	0.66	0.11	
1057	urn	turn	.9rn	t9rn	proverb one good turn deserves another	240	B			NN1	NN1	S+	0.07	0.0175	
1058	vassal	vessel	v&s0l	vEs0l		243	B			NN1	NN1		0.19	0.038	
1059	vicar	liquor	vIk0r	lIk0r	proverb candy is dandy, but liquor is quicker	245	B			NN1	NN1	S+	0.44	0.088	
1060	vigil	Virgil	vI.JIl	v9rJIl		245	B			NN1	NP0		0.33	0.055	
1061	word	work	w9rd	w9rk	proverb a woman's work is never done	255	B			NN1	NN1	S+	0.52	0.13	
1062	word	bird	w9rd	b9rd	proverb a bird in the hand ...	255	B			NN1	NN1	S+	0.66	0.165	
1063	yule	fool	jul	ful	proverb there's no fool like an old fool	261	B			NN1	NN1	S+	1	0.33333334	
1064	God I	gotta	ga?1	ga?0	idiom if you gotta go ...	99	T	atrocious, elliptical		NN1 PNP	VM0 TO0	S+	0.29	0.0725	
1065	abbreviate	a brie we ate	0brivi4t	0briwi4t		5	ne			VVB	AT0 NN1 PNP VVD		0.08	0.01	
1066	affiliate	a fill he ate	0fIl.i4t	0flIhi4t		9	ne			VVB	AT0 NN1 PNP VVD		0.07	0.00875	
1067	afflict	a flick	0flIkt	0flIk.	_def.	9	ne			VVB	ATO NN1		0.07	0.011666667	
1068	ain't	aunt	4nt	&nt	_def.	9	ne			VBB XX0	NN1		0.27	0.09	
1069	air	err	7r	9r	proverb to err is human, ... forgive, divine	92	B	see forklift		VVI	VVI	S+	0.3	0.15	
1070	amends	a men's	0mEndz	0mEn.z	_def.	12	ne			VVZ	AT0 NN1 POS		0.11	0.018333333	
1071	bait	beat	b4t	bit	idiom you can't beat it	22	B			VVB	VVB	E+	0.28	0.093333334	
1072	bead	beat	bid	bit	coll. beat it	25	B			VVI	VVI	I	0.09	0.03	
1073	beer	be	b6r	bi.	quote to be or not to be	27	B	beer (v.) is pushing it,		VVI	VBI	Q+	0.32	0.10666667	
1074	beguile	big girl	bIg1.l	bIg9rl	idiom she's a big girl now	27	B			VVB	AJ0 NN1	E+	0.33	0.055	
1075	bend	end	bEnd	.End	idiom the be-all and end-all	28	B			VVB	VVB	E+	0.21	0.0525	
1076	bind	mind	b1nd	m1nd	idiom mind over matter	30	ne	desp. WC B□,3		VVB	NN1	E+	0.44	0.11	
1077	bisex	bisect	b1sEks	b1sEkt	coll. bisecting angles	30	B			VVG	VVG	C	0.34	0.05666667	
1078	bleach	breach	bliC	briC	idiom breach of promise	31	B			VVI	VVI	I+	0.11	0.0275	
1079	blouse	blows	bl2z	bl5z		31	B			VVZ	VVZ		0.29	0.0725	
1080	braise	praise	br4z	pr4z	quote praise Him who made it	34	B	what?		VVB	VVB	Q+	0.14	0.035	
1081	bullet	pull it	bUlIt	pUlIt		38	B	Pun intr./Targ. trans. + O		VVB	VVB PNP		0.14	0.028	
1082	cauterize	caught a rise	kO?0r1z	ga?0r1z		44	B		Cockney	VVB	VVB AT0 NN1		0.36	0.05142857	
1083	cavort	quart	k0vOrt	k.wOrt	_def.	44	B		Swedish	VVI	NN1		0.12	0.02	+1
1084	chant	shan't	C&nt	S&nt		46	T	WS had chant-can't		VVB	VM0 XX0		0.19	0.0475	
1085	cost	lost	kast	last	title love's labour lost	56	B			VVN	AJ0	Q+	0.66	0.165	
1086	craze	praise	kr4z	pr4z	title praise the lord and pass the ammunition	57	B	spoon w/parse-pass		VVI	VVI	Q+	0.23	0.0575	
1087	crow	go	kr5	g.5	idiom if you gotta go...	59	B			VVI	VVI	S+	0.14	0.046666667	
1088	crunch	lunch	kr^nC	.l^nC	idiom out to lunch	59	B			VVI	NN1	I+	0.23	0.046	
1089	cull	call	k^l	kOl	idiom don't call us, we'll call you	60	B			VVB	VVB	E+	0.28	0.093333334	
1090	depart	the part	dIpart	D0part		65	T			VVB	AT0 NN1		0.06	0.01	
1091	deploy	the ploy	dIpl3	D0pl3		65	B			VVB	AT0 NN1		0.06	0.012	
1092	dry	cry	dr1	kr1	idiom see a grown man cry	71	B			AJ0	VVB	E+	0.52	0.17333333	
1093	earn	learn	.9rn	l9rn	coll. live and learn	73	B			VVB	VVB	B	0.08	0.02	
1094	eat	beat	.it	bit	idiom if you can't beat them, ...	74	B			VVI	VVI	S+	0.21	0.07	
1095	elate	be late	il4t	.l4t	idiom better late than never	75	B	WS elate-late		VVB	VBB AJ0	S+	0.13	0.0325	
1096	elevate	I'll have 8	El0v4t	1l0v4t		76	ne			VVB	PNP VM0 PN1		0.28	0.046666667	
1097	eloping	loping	il5pIN	.l5pIN	idiom go loping down the road	76	B			VVG	VVG	E+	0.13	0.021666666	+1

1098	enclose	in clothes	Enkl5.z	Inkl5Dz		77	B	ellipsis, complicated	VVB	PRP NN2		0.34	0.04857143	
1099	endow	indeed	End2.	Indid		77	ne		VVB	AV0		0.6	0.12	
1100	evidence	have a dance	.EvId0ns	h&v0d&ns	idiom may I have a dance with you	80	B		VVB	VHB AT0 NN1	I+	0.52	0.065	
1101	field	feel	fild	fil.	proverb you're only as young as you feel	86	B		VVB	VVB	S+	0.11	0.0275	
1102	flail	fail	fl4l	f.4l	idiom words fail me	88	B		VVB	VVB	E+	0.06	0.015	
1103	foist	first	f3.st	f9rst		90	T		VVB	AV0		0.33	0.066	
1104	forklift	forgive	fOrklIft	f0rg.Iv.	proverb to err is human, ...	92	B		VVI	VVI	S+	0.65	0.08125	
1105	fry	try	fr1	tr1	idiom try harder	94	B		VVB	VVB	I+	0.52	0.17333333	
1106	grow	glow	gr5	gl5	idiom glow with pride	103	B		VVI	VVI	E+	0.13	0.043333333	
1107	grow	blow	gr5	bl5	coll. blow your mind	103	B		VVB	VVB	I	0.44	0.14666666	
1108	herd	hurt	h9rd	h9rt	idiom you always hurt the one you love	110	B		VVI	VVI	E+	0.09	0.0225	
1109	hew	you	hju	.ju		110	T	what?	VVI	PNP		0.08	0.026666667	
1110	immerse	worse	Im9rs	.w9rs	idiom from bad to worse	117	T		VVI	AJC	E+	1.1	0.22	+1
1111	immolate	him more late	.Im0.l4t	hImOrl4t	foreigner talk	117	T		VVB	PNP AJC		0.36	0.045	
1112	impale	in pail	Imp4l	Inp4l	coll. pail bucket	118	T		VVB	PRP NN1	R	0.11	0.022	
1113	inhale	in hell	Inh4l	InhEl		119	B		VVI	PRP NN1		0.26	0.052	
1114	inhibit	inhabit	InhIbIt	Inh&bIt	quote they shall inhabit the earth	119	B		VVI	VVI	Q+	0.28	0.04	
1115	inoculate	immaculate	Inakjul4t	Im&kjulIt	coll. immaculate conception	120	T		VVB	AJ0	C	0.6	0.06666667	
1116	insert	assert	Ins9rt	0.s9rt	idiom assert one's masculinity	120	B		VVI	VVI	I+	0.11	0.018333333	e
1117	jab	'd you have	J.&b	J0&v		123	B		VVB	VDD PNP VHI		0.25	0.0625	-1
1118	joke	choke	J5k	C5k	coll. choke to death	125	B		VVI	VVI	I	0.18	0.06	
1119	lacquer	like her	l&k0r	l1k0r		133	B		VVB	VVI PNP		0.22	0.044	
1120	lath	laugh	l&T	l&f	proverb he who laughs last ...	134	B		VVZ	VVZ	S+	0.44	0.14666666	
1121	lease	leave	lis	liv	idiom love it or leave it	135	B		VVB	VVB	S+	0.66	0.22	
1122	let	met	lEt	mEt	idiom never met a man I didn't like	137	T		VMN	VVN	I+	0.38	0.12666667	
1123	lop	top	lap	tap	coll. top banana	142	T		VVI	AJ0	C	0.38	0.12666667	
1124	maimed	named	m4md	n4md	title a streetcar named desire	145	ne		VVN	VVN	Q+	0.11	0.0275	
1125	neap	reap	nip	rip	quote as you sow, so shall ye reap	158	B		VVB	VVI	Q+	0.29	0.096666664	
1126	needle	need all	nid0l	nidOl		158	B		VVB	VVB AT0		0.25	0.05	
1127	nod	not	nad	nat		160	B	morpheme, ok when part of word	NN1	XX0		0.09	0.03	
1128	paint	ain't	p4nt	.4nt	idiom ... isn't what it used to be	169	B		VVB	VBB XX0	E+	0.16	0.04	
1129	park	bark	park	bark	idiom his bark is worse than his bite	170	T		NN1	NN1	E+	0.16	0.04	
1130	parse	pass	pars	p&.s	title praise the lord and pass the ammunition	8	B	spoon w/craze-praise	VVB	VVB	Q+	0.21	0.0525	
1131	pawn	upon	.pOn	0pan	quote return with your shields ...	172	B		VVB	PRP	Q+	0.32	0.08	
1132	ponder	yonder	pand0r	jand0r	coll. wide blue yonder	179	B		VVB	VVB	C	1	0.16666667	
1133	punch	lunch	p^nC	l^nC	coll. out to lunch	185	B		NN1	NN1	I	1	0.25	
1134	raise	praise	.r4z	pr4z	quote ...o bury Caesar, not to praise him	191	B		VVI	VVI	Q+	0.24	0.06	
1135	rate	ate	r4t	.4t	idiom something that I ate	192	B		VVD	VVD	I+	0.03	0.01	
1136	razor	raise your	r4z.0r	r4zj0r	idiom raise eyebrows	192	B		VVI	VVI DPS	I+	0.09	0.015	
1137	reap	weep	rip	wip	idiom read it and weep	192	B		VVB	VVB	I+	0.21	0.07	
1138	reek	meek	rik	mik	quote the meek shall inherit the earth	194	T		VVB	VV0	Q+	0.66	0.22	
1139	relapse	relax	rIl&ps	rIl&ks	coll. relax and enjoy	194	B		VVI	VVI	B	0.24	0.04	
1140	repeat	repent	rIpi.t	rIpEnt	proverb marry in haste, repent in leisure	195	B		VVB	VVB	S+	0.38	0.06333333	
1141	roil	rule	r3l	rul	quote Britain rule the waves	198	B		VVZ	VVZ	Q+	0.3	0.1	
1142	ruminate	Roman ate	rum0n4t	r5m0n4t		200	B		VVB	NP0 VVD		0.27	0.03857143	
1143	seek	sheik	sik	Sik	idiom sheik of Araby	206	T		VVB	NN1	I+	0.17	0.05666667	
1144	sicken	thicken	sIk0n	TIk0n	idiom the plot thickens	210	B		VVZ	VVZ	I+	0.25	0.05	
1145	skid	kid	skId	.kId	idiom I kid you not	211	B		VVB	VVB	I+	0.08	0.02	
1146	slake	take	sl4k	t.4k	quote take five	212	B		VVI	VVI	Q+	0.4	0.1	
1147	slink	sink	slINk	s.INk	idiom sink or swim	213	B		VVI	VVI	B+	0.06	0.012	
1148	snow	snore	sn5.	snOr	_def.	214	ne		VVI	VVI		0.3	0.075	
1149	stink	think	stINk	T.INk	proverb great minds think alike	220	B		VVB	VVB	S+	0.32	0.064	
1150	stray	stay	str4	st.4	idiom something is here to stay	221	B		VVI	VVI	E+	0.03	0.0075	
1151	strike	like	str1k	..l1k	idiom I've never met a man I couldn't like	221	B		VVI	VVI	Q+	0.28	0.056	
1152	tack	take	t&k	t4k	idiom I just can't take it anymore	225	B		VVI	VVI	I+	0.27	0.09	
1153	sing	Thing	TIN	sIN	idiom it don't mean a thing ...	211	T		VVI	NN1	E+	0.11	0.036666665	
1154	thumb	sum	T^m	s^m	idiom sum up a case	230	B		VVI	VVI	I+	0.11	0.036666665	
1155	tinkle	twinkle	t.INk0l	twINk0l	title twinkle, twinkle, little star	232	B		VVB	VVB	Q+	0.13	0.018571429	
1156	tramp	cramp	tr&mp	kr&mp	idiom cramp one's style	235	B		VVI	VVI	E+	0.16	0.032	
1157	tryst	trust	trIst	tr^st	idiom in God we trust	236	B	wtf?	VVB	VVB	Q+	0.27	0.054	
1158	underwrite	on the right	^nd0r1t	anD0r1t	idiom on the right side of the ledger	240	T		VV0	PRP AT0 AJ0	E+	0.31	0.044285715	
1159	undulate	on the late	^ndj0l4t	0nD.0l4t		240	T		VV0	PRP AT0 AJ0		0.43	0.05375	e

1160	ventilate	went till 8	vEn?0l4t	wEntIl4t		244	T			VV0	VVD PRP NN1		0.29	0.03625		
1161	venture	went your	vEnC0r	wEn?0.	idiom go one's way	244	T			VV0	VVD DPS	E+	0.77	0.12833333		
1162	wand	want	wand	want	coll. anything you want	249	B			VVB	VVB	I	0.09	0.0225		
1163	wane	rain	w4n	r4n	idiom it never rains but it pours	249	B			VVZ	VVZ	E+	0.14	0.046666667		
1164	west	rest	wEst	rEst		251	B		rabbit	VVB	VVB		0.14	0.035		
1165	whine	shine	w1n	S1n	idiom rise and shine	252	B			VVB	VVB	B+	1	0.33333334		
1166	wilt	will	wIlt	wIl.	coll. I will	253	B			VVB	VMD	I	0.07	0.0175		
1167	worship	war-ship	w0rSIp	wOrSIp		255	T			NN1	NN1		0.25	0.041666668		
1168	yawn	you're on	j..On	j8ran	coll. you're on	259	B	P: exclamation		VV0	PNP VBB AVP	P	1.29	0.258	e	-1
1169	yearn	earn	j9rn	.9rn	idiom earn one's way	259	B			VVI	VVI	I+	0.1	0.025		
1170	yonder	(a)re on their	jand0r	ranD0r	idiom on one's way home	260	T			PRP	VBB PRP DPS	I+	0.71	0.11833333		
1171	zing	sing	zIN	sIN		262	ne	what?		VV0	VV0		0.07	0.023333333		
1172	mend straight	menstruate	mEndstr4t	mEn.str4t		148	B			VVI AVP	VVI		0.11	0.012222222		
1173	share an aid	serenade	S7r0n4d	sEr0n4d	quote serenade ... 'neath a fair lady's window	254	B	see windrow		VVB AT0 NN1	VVI	Q+	0.38	0.054285713		
1174	feed 'em	freedom	f.id0m	frid0m	idiom battle cry of freedom	85	T			VVB PNP	NN1	Q+	0.04	0.006666667		
1175	trust Owens	throw stones	tr^st50nz	Tr5st5.nz	proverb people who live in glass houses...	113	B	see hose		VVI NP0	VVI NN2	S+	0.51	0.05666667		
1176	watch her	water	waC0r	wa?0r	coll. Scotch and water	250	B			VVI PNP	NN1	B	0.66	0.132		
1177	wore cologne	walk alone	wOrk0l5n	wO.k0l5n		52	B	see cologne		VVB	VVB		0.03	0.00375		
1178	blown apart	Bonaparte	bl5n0part	b.5n0part	coll. Napoleon Bonaparte	31	B			VVB AV0	NP0	I	0.06	0.006666667		
1179	ginned	sinned	JInd	sInd	idiom forgive me, father, for I have sinned	97	B			VVN	VVN	Q+	0.38	0.095		
1180	invading	evading	Inv4dIN	i.v4dIN	idiom evade the issue	121	B			VVG	VVG	I+	0.25	0.035714287		
1181	open	hopin'	.5p0n	h5pIn	idiom keep hoping	165	B			AV0	VVG	I+	0.07	0.014		
1182	zebra	the bra	zibr0	D0bra		261	T			NN1	AT0 NN1		0.84	0.168	e	

Appendix F: Intrusion Frequency Table

The following table is taken from Sobkowiak (1991: 223). It is the primary basis for the cost table in Appendix G.

T\P	S	Z	h	m	n	N	l	r	j	w	?	i	I	E	&	a	O	U	u	9	0	^	1	2	3	4	5	6	7	8	.	sum
f	1	0	5	0	1	0	1	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	67
v	0	0	0	3	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	81
T	2	0	0	0	1	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	61
D	1	0	1	0	1	0	3	2	1	1	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	237
s	52	0	3	1	1	0	2	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	265
z	7	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	134
S		2	2	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	50
Z	4		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
h	2	0		5	3	0	3	2	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	132
m	0	0	3		28	0	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	78
n	0	0	3	38		16	7	7	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	134
N	0	0	0	3	26		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	31
l	1	0	1	0	12	0		30	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63	141
r	0	0	0	2	8	0	40		2	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	224	334
j	0	0	0	0	0	0	1	5		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57	83
w	0	0	2	1	3	0	4	14	1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	138
?	0	0	0	0	1	0	0	4	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	78
i	0	0	0	0	0	0	0	0	0	0	0		79	5	2	1	1	0	1	0	20	0	4	1	0	13	0	5	0	0	8	140
I	0	0	0	0	0	0	0	0	0	0	0	106		50	20	10	2	2	4	1	161	6	9	0	0	11	3	3	1	0	40	429
E	0	0	0	0	0	0	0	0	0	0	0	12	55		57	18	2	1	2	4	37	24	2	1	0	20	3	3	5	0	1	247
&	0	0	0	0	0	0	0	0	0	0	0	1	13	51		63	8	0	2	5	31	12	3	2	0	17	1	0	3	0	0	212
a	0	0	0	0	0	0	0	0	0	0	0	3	10	6	37		29	3	4	8	19	24	0	1	1	3	14	1	4	0	2	169
O	0	0	0	0	0	0	0	0	0	0	0	0	0	6	17	33		4	5	3	26	10	0	4	3	3	20	1	0	5	0	140
U	0	0	0	0	0	0	0	0	0	0	0	0	6	1	1	0	1		8	0	10	10	0	0	0	0	4	0	0	1	1	43
u	0	0	0	0	0	0	0	0	0	0	0	1	4	0	1	2	3	2		0	10	3	0	3	2	0	15	0	0	3	0	49
9	0	0	0	0	0	0	0	0	0	0	0	0	5	4	6	7	7	0	2		4	3	3	1	7	0	2	2	1	0	0	54
0	0	0	0	0	0	0	0	0	0	0	0	31	154	24	35	34	17	5	20	5		10	5	2	1	9	40	2	4	1	77	476
^	0	0	0	0	0	0	0	0	0	0	0	0	14	8	15	48	13	4	10	2	16		0	0	0	2	12	0	0	0	0	144
1	0	0	0	0	0	0	0	0	0	0	0	10	13	12	44	36	4	0	1	2	51	4		3	10	19	7	0	1	0	2	219
2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	7	15	2	0	3	0	6	2	2		0	0	3	0	0	1	0	42
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	2	0	0	3	0		0	1	0	0	0	0	9
4	0	0	0	0	0	0	0	0	0	0	0	15	10	12	16	5	0	0	1	1	13	1	14	0	4		2	0	0	0	0	94
5	0	0	0	0	0	0	0	0	0	0	0	2	5	3	3	25	10	2	14	5	28	7	2	4	6	2		0	0	0	2	120
6	0	0	0	0	0	0	0	0	0	0	0	4	2	4	0	1	1	0	0	1	3	0	1	0	0	3	0		6	0	1	27
7	0	0	0	0	0	0	0	0	0	0	0	2	1	11	22	8	5	0	0	5	6	0	0	0	0	2	0	3		0	1	66
8	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	4	4	5	5	4	2	0	0	0	0	0	2	0	0		1	30
.	2	0	74	25	47	5	104	280	45	39	5	29	50	5	3	4	3	0	6	1	203	1	1	0	0	2	11	0	0	0		1264
sum	103	15	105	87	152	21	189	368	70	93	54	216	422	205	286	316	113	28	88	49	646	117	49	22	34	106	140	20	25	11	958	

Appendix G: Cost Table

T\P	p	t	k	b	d	g	C	J	f	v	T	D	s	z	S	Z	h	m	n	N	l	r	j	w	?
p	0.00	0.31	0.23	0.14	1.00	1.00	1.00	1.00	0.23	0.66	1.00	1.00	0.66	1.00	1.00	1.00	1.00	0.66	0.66	1.00	1.00	0.66	1.00	1.00	0.66
t	0.18	0.00	0.14	0.44	0.09	1.00	0.38	1.00	0.52	0.66	0.20	1.00	0.34	0.52	1.00	1.00	1.00	1.00	0.34	1.00	0.38	0.34	1.00	1.00	0.21
k	0.24	0.16	0.00	1.00	0.52	0.10	0.38	1.00	0.44	1.00	1.00	1.00	0.34	1.00	1.00	1.00	0.34	1.00	1.00	1.00	0.66	1.00	1.00	1.00	0.66
b	0.16	0.38	0.52	0.00	0.38	0.31	1.00	1.00	0.38	0.44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.66	0.66	1.00	0.66	0.66	1.00	0.66	1.00
d	0.44	0.11	0.34	0.21	0.00	0.44	1.00	0.66	1.00	0.44	0.66	0.44	0.66	0.66	1.00	1.00	1.00	0.66	0.38	1.00	0.34	0.44	1.00	1.00	0.18
g	0.66	0.66	0.11	0.44	0.34	0.00	1.00	0.44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	0.38	0.52	1.00	1.00	1.00	0.00	0.18	1.00	1.00	1.00	1.00	1.00	1.00	0.15	1.00	1.00	1.00	0.66	1.00	1.00	1.00	1.00	1.00	1.00
J	1.00	1.00	1.00	0.66	0.44	0.66	0.20	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.38	0.31	1.00	1.00	1.00	1.00	1.00	1.00	0.44	1.00	1.00
f	0.21	1.00	1.00	0.66	1.00	0.52	1.00	1.00	0.00	0.17	0.44	1.00	0.34	1.00	1.00	1.00	0.38	1.00	1.00	1.00	1.00	0.52	1.00	1.00	1.00
v	1.00	0.66	1.00	0.18	0.66	0.66	1.00	1.00	0.14	0.00	1.00	1.00	0.66	0.52	1.00	1.00	1.00	0.52	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T	0.52	0.18	1.00	1.00	0.52	1.00	1.00	0.66	0.24	1.00	0.00	0.66	0.25	0.66	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.66
D	1.00	0.18	0.66	1.00	0.05	0.44	1.00	1.00	1.00	0.38	0.23	0.00	0.38	0.34	1.00	1.00	1.00	1.00	1.00	1.00	0.52	0.66	1.00	1.00	0.23
s	1.00	0.44	0.27	0.66	1.00	1.00	0.44	0.38	0.27	1.00	0.11	1.00	0.00	0.07	0.09	1.00	0.52	1.00	1.00	1.00	0.66	0.66	1.00	1.00	1.00
z	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.44	0.44	0.07	0.00	0.31	0.38	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
S	1.00	1.00	1.00	1.00	1.00	1.00	0.19	0.52	1.00	1.00	1.00	1.00	0.17	1.00	0.00	0.66	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Z	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.44	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h	1.00	0.44	0.38	0.52	1.00	1.00	0.52	1.00	0.52	0.52	1.00	1.00	0.44	1.00	0.66	1.00	0.00	0.38	0.52	1.00	0.52	0.66	0.44	0.38	1.00
m	0.52	1.00	0.66	0.44	0.44	0.66	1.00	1.00	0.44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52	0.00	0.14	1.00	0.38	0.66	1.00	1.00	1.00
n	1.00	0.66	0.66	0.52	1.00	1.00	1.00	1.00	0.66	0.66	1.00	1.00	1.00	1.00	1.00	1.00	0.52	0.11	0.00	0.19	0.31	0.31	1.00	0.52	1.00
N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52	0.14	0.00	1.00	1.00	1.00	1.00	1.00
l	1.00	1.00	0.66	1.00	0.44	1.00	1.00	1.00	1.00	0.44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.23	1.00	0.00	0.13	0.25	0.25	1.00
r	0.52	0.52	1.00	0.66	0.29	0.52	1.00	0.38	1.00	0.44	1.00	1.00	0.44	1.00	1.00	1.00	1.00	0.66	0.29	1.00	0.11	0.00	0.66	0.14	1.00
j	1.00	1.00	1.00	0.66	1.00	1.00	1.00	0.27	1.00	1.00	1.00	1.00	1.00	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.38	0.00	0.66	1.00
w	1.00	1.00	1.00	0.52	1.00	0.66	1.00	0.66	0.34	0.08	1.00	1.00	1.00	1.00	1.00	1.00	0.66	1.00	0.52	1.00	0.44	0.21	1.00	0.00	1.00
?	1.00	0.17	0.66	1.00	0.12	0.66	0.66	1.00	0.66	1.00	1.00	1.00	0.52	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.44	1.00	1.00	0.00
i	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
&	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
O	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
U	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
u	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
^	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
.	0.16	0.07	0.10	0.21	0.11	0.18	1.00	0.34	0.27	0.34	0.38	0.66	0.08	0.25	0.66	1.00	0.08	0.14	0.10	0.38	0.06	0.03	0.10	0.11	0.38

T\P	i	I	E	&	a	O	U	u	9	0	^	1	2	3	4	5	6	7	8	.
p	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.24
t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.07
k	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.18
b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.21
d	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.09
g	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.38
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
J	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52
f	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.44
v	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.17
T	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.13
s	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.10
z	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.21
S	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52
Z	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.07
m	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.19
n	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.10
N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
l	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08
r	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.04
j	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.09
w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.13
?	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.34
i	0.00	0.15	0.29	0.30	0.30	0.30	0.40	0.30	0.40	0.26	0.40	0.29	0.30	0.40	0.28	0.40	0.29	0.40	0.40	0.29
I	0.10	0.00	0.21	0.26	0.28	0.30	0.30	0.29	0.30	0.00	0.29	0.28	0.40	0.40	0.28	0.29	0.29	0.30	0.40	0.11
E	0.28	0.20	0.00	0.19	0.27	0.30	0.30	0.30	0.29	0.23	0.26	0.30	0.30	0.40	0.26	0.29	0.29	0.29	0.40	1.00
&	0.30	0.28	0.20	0.00	0.18	0.29	0.40	0.30	0.29	0.24	0.28	0.29	0.30	0.40	0.27	0.30	0.40	0.29	0.40	1.00
a	0.29	0.28	0.29	0.23	0.00	0.25	0.29	0.29	0.29	0.26	0.26	0.40	0.30	0.30	0.29	0.27	0.30	0.29	0.40	0.66
O	0.40	0.40	0.29	0.27	0.24	0.00	0.29	0.29	0.29	0.25	0.28	0.40	0.29	0.29	0.29	0.26	0.30	0.40	0.29	1.00
U	0.40	0.29	0.30	0.30	0.40	0.30	0.00	0.29	0.40	0.28	0.28	0.40	0.40	0.40	0.40	0.29	0.40	0.40	0.30	1.00
u	0.30	0.29	0.40	0.30	0.30	0.29	0.30	0.00	0.40	0.28	0.29	0.40	0.29	0.30	0.40	0.27	0.40	0.40	0.29	1.00
9	0.40	0.29	0.29	0.29	0.29	0.29	0.40	0.30	0.00	0.29	0.29	0.29	0.30	0.29	0.40	0.30	0.30	0.30	0.40	1.00
0	0.24	0.01	0.26	0.23	0.24	0.27	0.29	0.26	0.29	0.00	0.28	0.29	0.30	0.30	0.28	0.23	0.30	0.29	0.30	0.07
^	0.40	0.27	0.29	0.27	0.21	0.28	0.29	0.28	0.30	0.27	0.00	0.40	0.40	0.40	0.30	0.28	0.40	0.40	0.40	1.00
1	0.28	0.28	0.28	0.22	0.23	0.29	0.40	0.30	0.30	0.20	0.29	0.00	0.29	0.28	0.26	0.29	0.40	0.30	0.40	0.66
2	0.40	0.30	0.40	0.29	0.27	0.30	0.40	0.29	0.40	0.29	0.30	0.30	0.00	0.40	0.40	0.29	0.40	0.40	0.30	1.00
3	0.40	0.40	0.40	0.40	0.30	0.30	0.40	0.40	0.30	0.40	0.40	0.29	0.40	0.00	0.40	0.30	0.40	0.40	0.40	1.00
4	0.27	0.28	0.28	0.27	0.29	0.40	0.40	0.30	0.30	0.28	0.30	0.27	0.40	0.29	0.00	0.30	0.40	0.40	0.40	1.00
5	0.30	0.29	0.29	0.29	0.25	0.28	0.30	0.27	0.29	0.25	0.29	0.30	0.29	0.29	0.30	0.00	0.40	0.40	0.40	0.66
6	0.29	0.30	0.29	0.40	0.30	0.30	0.40	0.40	0.30	0.29	0.40	0.30	0.40	0.40	0.29	0.40	0.00	0.29	0.40	1.00
7	0.30	0.30	0.28	0.26	0.29	0.29	0.40	0.40	0.29	0.29	0.40	0.40	0.40	0.40	0.30	0.40	0.29	0.00	0.40	1.00
8	0.40	0.40	0.29	0.40	0.29	0.29	0.29	0.29	0.29	0.30	0.40	0.40	0.40	0.40	0.40	0.30	0.40	0.40	0.00	1.00
.	0.13	0.10	0.38	0.52	0.44	0.52	1.00	0.34	1.00	0.04	1.00	1.00	1.00	1.00	0.66	0.24	1.00	1.00	1.00	0.00

This preliminary cost table is based on the following formulae where x is frequency attested by Sobkowiak as documented in Appendix F and 161 is the highest frequency among Vs:

segment pair type	formula
C/C C/. V/.:	$y=x^{-0.6}$ (asymptotic)
V/V:	$y=0.3-0.3x/161$ (linear)
C/V & V/C:	1 (unattested)

VITA

Christian F. Hempelmann, born 5 February 1972 in Hannover, Germany.

Education

- Ph.D. in Linguistics
  student since 2000 (candidate since July 2002, successful defense 28 April 2003, graduation 8 August 2003): Interdepartmental Program in Linguistics at Purdue University,
  - dissertation "Paronomasic Puns: Target Recoverability towards Automatic Generation"
  - dissertation committee: Victor Raskin (chair), Salvatore Attardo, Shaun Hughes, Mary Niepokuj, Sergei Nirenburg
  - primary area: computational linguistics (ontological semantics, NLP for information security); secondary areas: linguistic humor studies, theoretical linguistics, and historical linguistics
- *M.A. in English*
  2000 specialization in TESOL, Youngstown State University. Thesis: "Incongruity and Resolution of Humorous Narratives-Linguistic Humor Theory and the Medieval Bawdry of Rabelais, Boccaccio, and Chaucer" supervised by Salvatore Attardo
- *M.A. in English Linguistics, secondary area Religious Studies*
  1998, [Magister Artium der Englischen Sprachwissenschaft, Zweitfach Religionswissenschaft], Universität Hannover, Thesis: "Linguistic Approaches to Humor: The General Theory of Verbal Humor in Performance" supervised by Rainer Schulze
- *Abitur*
  1991, major in Ancient Greek, German literature, and mathematics, Kaiser-Wilhelms-Gymnasium, Hannover, Germany

Research and Professional Experience

- *since Aug. 2000*
  research assistant in various NLP projects with Victor Raskin and Mikhail Atallah at the Center for Education and Research in Information Assurance and Security (CERIAS) at Purdue University

- *since Aug. 2000*
  teaching assistant in the English Department of Purdue University: taught introductory composition and linguistics courses
- *Aug. 1998 - June 2000*
  teaching assistant in the English Department of Youngstown State University: taught introductory composition and linguistics courses; helped with German courses in the Foreign Languages Department
- *Oct. 1999 - May 2000*
  created the camera-ready copy for a linguistics textbook using the LaTeX system, including creation of the graphics, writing part of the glossary, and indexing the book: Brown, Steven and Salvatore Attardo. *Understanding Language - Structure, Interaction, and Variation*. Ann Arbor: University of Michigan Press. 2000.
- *Aug. 1991 - Aug. 1998*
  employed by Textilwerkstatt-Verlag, European Textile Network: translation of German texts into English and other editorial and DTP work for the international quarterly *Textileforum*, its web site, and other publications, supervising the computer systems and network (Apple Macintosh) of the publishing house, responsible for the subscriber database of the quarterly, working with the following programs: ACI 4th Dimension (including programming of new processes for various databases), QuarkXpress, Aldus PageMaker, Adobe PhotoShop, Microsoft Word, Excel, PowerPoint, etc.
- *Nov. 1997 - Mar. 1998*
  created the print-ready copy of a linguistics anthology in QuarkXpress: Schulze, Rainer. ed. *Making Meaningful Choices in English*. Tübingen: Gunter Narr. 1998.
- *July 1995 - Sept. 1995*
  internship at the University of Central England in Birmingham, UK

Teaching Experience

- *Introduction to Linguistics*
  YSU, Spring 1999 to Fall 2000; introductory undergraduate linguistics course
- *Composition I*
  YSU, Fall 1998 to Spring 1999; Purdue, Fall 2000; freshman composition course
- *Elements of Linguistics*
  Purdue, Fall 2002; introductory undergraduate linguistics course
- *Composition II*
  Purdue, Spring 2001; sophomore composition course
- *Introduction to Linguistics*
  Purdue, Fall 2001, Summer 2003; supplementary teaching for Victor Raskin in graduate-level introduction course
- *Ontological Semantics*
  Purdue, Fall 2001; supplementary teaching for Victor Raskin in graduate-level computational linguistics course

Academic and University Service

- *since 2001*
  Member of the dissertation committee of Michele Sala, Università degli Studi di Bergamo, Italy
- *since 2000*
  Referee for HUMOR - International Journal of Humor Research
- *since 2000*
  Member of the Computers and Composition Committee, Purdue Macintosh Tutorials and Resource Collections
- *since 2002*
  Secretary of the Purdue Linguistics Association (PLA)
- *2000 – 2002*
  Founding President of the Purdue Linguistics Association (PLA)
- *July 2002*
  Reviewer for the U.S. Civilian Research & Development Foundation for the Independent States of the Former Soviet Union (CRDF)
- *1999 – 2000*
  Graduate Student Delegate to the English Department, YSU
- *2000*
  German Tutor in the Foreign Language Department, YSU

Membership in Professional Societies

- Linguistic Society of America
- International Society for Humor Studies
- Medieval Association of the Midwest
- Modern Language Association

Publications

- Book Editorship
  - *Proceedings of the New Security Paradigms Workshop 2002.* with Victor Raskin. Eds. New York: ACM. 2003.
  - *Proceedings of the New Security Paradigms Workshop 2001.* with Victor Raskin. Eds. New York: ACM. 2002.
- Papers
  - " '99 nuns giggle, 1 nun gasps:' The not-all-that-Christian Natural Class of Christian Jokes." *HUMOR - International Journal of Humor Research.* 16-1. 2003: 1-31.
  - "Natural Language Watermarking and Tamperproofing." with Mikhail J. Atallah, Victor Raskin, Mercan Karahan, Radu Sion, Umut Topkara, and Katrina E. Triezenberg. In: Fabien A. P. Petitcolas. Ed. *Information Hiding, 5th International Workshop, IH 2002.* Berlin: Springer. 2002: 196-210.

- "Script oppositions and logical mechanisms: Modeling incongruities and their resolutions." with Salvatore Attardo and Sara DiMaio. *HUMOR - International Journal of Humor Research* 15-1. 2002: 3-46.
- "Why NLP should move into IAS." with Victor Raskin, Sergei Nirenburg, Mikhail J. Atallah, and Katrina E. Triezenberg. In: Krauwer, Steven (on behalf of ELSNET). Ed. *Proceedings of the Workshop on a Roadmap for Computational Linguistics at COLING 2002, Taipei, August 31, 2002.* Utrecht: ELSNET. 2002: 1-7.
- "Ontology in Information Security: A Useful Theoretical Foundation and Methodological Tool." with Victor Raskin, Katrina E. Triezenberg, and Sergei Nirenburg. In: Raskin, Victor and Christian F. Hempelmann. Eds. *Proceedings of the New Security Paradigms Workshop 2001.* New York: ACM. 2002: 53-59.
- "Natural Language Watermarking: Design, Analysis, and a Proof-of-concept Implementation." with Victor Raskin, Mikhail Atallah, Florian Kerschbaum, Sanket Naik, Michael Crogan, and Dina Mohamed. In: Moskowitz, Ira S. Ed. *Information Hiding. Proceedings of the Fourth International Workshop IH 2001.* Berlin: Springer: 185-199. 2001.

- Reviews
  - "Uwe Wirth: *Diskursive Dummheit: Abduktion und Komik als Grenzphänomene des Verstehens.* [Stupidity in Discourse: Abduction and Humor as Fringe Phenomena of Understanding.] Heidelberg: Winter. 1999." *HUMOR - International Journal of Humor Research*. Forthcoming. 16-3. 2003.

Presentations

- Conference Presentations
  - "The Genesis of a Script for Bankruptcy in Ontological Semantics." with Victor Raskin, Sergei Nirenburg, Inna Nirenburg, and Katrina E. Triezenberg. at the Text Meaning Workshop of the HLT-NAACL 2003 Human Language Technology Conference in Edmonton, Canada, May 31, 2003.
  - "YPS-The Ynperfect Pun Selector for Computational Humor." at the CHI2003 of the ACM, Ft. Lauderdale, FL, April 6, 2003.
  - "Permissible Distance and Recoverability in Paronomasic Puns." at the 14th ISHS Conference, University of Bologna, Bertinoro, Italy, July 5, 2002.
  - " '99 nuns giggle, 1 nun gasps:' The not-all-that-Christian Natural Class of Christian Jokes" at the 2001 Conference of the International Society for Humor Studies in College Park, MD, July 7, 2001.
  - "Developing an Automatic Hybrid Data and Text System for Downgrading Sensitive Documents (Poster)." with Victor Raskin, Mikhail Atallah and Dina Mohamed. at the Second Annual Research Symposium of the Center for Education and Research in Information Assurance and Security (CERIAS), Purdue University, West Lafayette, IN, April 26, 2001.
  - "Watermarking of Natural Language Text (Poster)." with Victor Raskin, Mikhail Atallah, Florian Kerschbaum, Sanket Naik, Michael Crogan, and Dina

Mohamed. at the Second Annual Research Symposium of the Center for Education and Research in Information Assurance and Security (CERIAS), April 26, 2001.

- "The Joking Monk: Humor and Narrative Structure in Rabelais' *Gargantua*" at the 1999 Conference of the International Society for Humor Studies in Oakland, CA, July 2, 1999.

- Invited Presentations
  - "YPS: The Ynperfect Pun Selector for Computational Humor." at Twente University, Enschede, The Netherlands, May 15, 2003.
  - "YPS: The Ynperfect Pun Selector for Computational Humor." at the ITC-irst Centro per la Ricerca Scientifica e Tecnologica, Trento, Italy, May 6, 2003.
  - "Natural Language Information Assurance and Security." with Victor Raskin. at the Center for Education and Research in Information Assurance and Security (CERIAS) Security Seminar, Purdue University, West Lafayette, IN, April 8, 2003.
  - "Current Issues in Linguistic Humor Research: An Overview." LinguA guest lecture at the Universität Hannover, Germany, June 18, 2002.
  - "Current Issues in Linguistic Humor Research: An Overview." at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, June 6, 2002.
  - "Ontology for IAS: Extending Research and Application Paradigms In Information Security." with Victor Raskin and Katrina E. Triezenberg. at the Center for Education and Research in Information Assurance and Security (CERIAS) Security Seminar, Purdue University, West Lafayette, IN, February 6, 2002.

Awards

- *2003*
  GSA travel award to support the presentation of research at the Computer-Human Interface (CHI) 2003 Conference of the Association for Computing Machinery (ACM) in Ft. Lauderdale, FL
- *2001*
  Quadrille Ball Scholarship for research in natural language processing and information security
- *2001*
  GSA travel award to support the presentation of research at the Conference of the International Society of Humor Studies in College Park, MD