2005

Topology, morphisms, and randomness in the space of formal languages

David E. Kephart
University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the American Studies Commons

Scholar Commons Citation
Kephart, David E., "Topology, morphisms, and randomness in the space of formal languages" (2005). Graduate Theses and Dissertations.
http://scholarcommons.usf.edu/etd/721

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Topology, Morphisms, and Randomness
in the Space of Formal Languages

by

David E. Kephart

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctorate of Philosophy
Department of Mathematics
College of Arts and Sciences
University of South Florida

Major Professor: Nataša Jonoska, Ph.D.
Gregory McColm, Ph.D.
Arunava Mukherjea, Ph.D.
Masahiko Saito, Ph.D.
Stephen Suen, Ph.D.

Date of Approval:
June 20, 2005

Keywords: entropy, language norm, language distance, language space, pseudo-metric, symbolic dynamics

© Copyright 2005, David E. Kephart
Lacking a reason, she said, why continue?
In her calm question I gained purpose
And now she requests I not mention her name!
On such irony, life turns.

Yes, I could name others, my brother Michael,
It would be fitting, but not complete; to
Her of long suffering and practical optimism, of
Up-turned face and crystalline realism——
I dedicate this to my wife, who gave me how and
why to go on.
Acknowledgments

Not one word (or symbol!) of this dissertation could be conceived without the constant, one might say unrelenting support of Natasha Jonoska. Her insight into language theory, her awareness of the examples, counter-examples and interconnections within the field, her anticipation of the thousand practical details that attend every scientific endeavor, her rigorous criticism, her refusal to settle for less than what can actually be obtained from a subject, not to mention her frequent and telling suggestions: all of these were vital to the production of this work. I can only hope to have in some degree merited such unstinting aid.

It has also been my great privilege to work with committee members of stature in the mathematical community. I thanks them, each and severally, for having generously contributed their time and skills (not to mention patience) to reviewing and critiquing this dissertation, despite myriad other demands upon their crowded schedules. It has been my good fortune to have studied subjects vital to the production of this dissertation with Greg McColm, Arunava Mukherjea, and Masahiko Saito. Stephen Suen, though he may not know it, inspired some of my first forays into problems of discrete mathematics. And I thank Dewey Rundus, chairman of my committee, for taking on that obligation without a qualm.

There are many other faculty members who have helped me and without whom I would have but a vague idea of what mathematics is all about. In particular, I thank Jogindar Ratti for his encouragement over the years, Brian Curtin for numerous enlightening discussions, Evgenii Rakhmanov for a comprehensive vision of complex analysis and approximation theory, and also Edwin Clark for bringing Normalized Google Distance to my attention. I have also incurred a deep debt to Scott Rimbe, who has rescued me on numerous occasions when health and my naiveté led me into conflicts with my teaching obligations.

The past five years at the University of South Florida have been remarkable because I have studied with a class of graduate students of truly imposing ability. There is not space for every name, but my special thanks goes to Edgardo Cureg both for his assistance with the calculations required in Chapter Six and for many other favors along the way, to Michiru Shibata, who has assisted me both with English and schedule conflicts, and to Luis Camara, the first student math speaker I heard on the campus. But I must give a special credit and appreciation to Kalpana Mahalingam, due to whose tenacity I first entered the remarkable world of language theory. Like her advisor, she knew what was best for me.

A final word of thanks to Aya Ossowski, who not only informed me of the due dates and format requirements, but assisted me enthusiastically in meeting them.

All shortcomings which remain in this document, and I am uncomfortably aware that they exist, may be credited to myself alone. They are to be regarded as errors to which I have somehow clung despite everything Natasha, and everyone else, has done on my behalf.

The research for this dissertation was sponsored by the National Science Foundation, through grants CCA #0432009 and EIA #0086015.
# Table of Contents

List of Tables ............................................. iii
List of Figures ........................................... iv
Abstract ................................................... v

Chapter 1  Formalizing distance and topology on a space of languages ............... 1
  1.1 Are languages the support of a meaningful distance? .......................... 1
  1.2 Basic definitions .......................................... 5
    1.2.1 General notation ...................................... 5
    1.2.2 Languages and language families .......................... 6
    1.2.3 Language spaces ....................................... 9
  1.3 Morphisms on words and language spaces ........................................ 10
    1.3.1 Extending literal mappings and word morphisms to language morphisms . 11
    1.3.2 Language Space Isomorphisms ................................ 20
    1.3.3 The permutative set difference of languages ................. 22
  1.4 Language pseudo-metrics and language norms .................................... 24
    1.4.1 Definition of language pseudo-metric and language norm ............. 25
    1.4.2 Link between language norms and language pseudo-metrics .......... 26
  1.5 Random languages ......................................... 29
    1.5.1 Martin-Löf randomness tests ................................ 31
    1.5.2 A general approach to randomness in topological language spaces .... 33

Chapter 2  Formal language space as a Cantor space .................................. 35
  2.1 The Cantor language norm and distance .......................................... 36
    2.1.1 Definition .............................................. 36
    2.1.2 Language cylinder sets and their enumeration ....................... 38
  2.2 The Cantor topology on a language space ....................................... 41
  2.3 Randomness under Cantor distance .............................................. 42
    2.3.1 An upper-semi-computable measure on open sets .................. 42
    2.3.2 Examples of nonrandom languages ................................ 46
    2.3.3 Random equals non-recursively enumerable .......................... 47

Chapter 3  The Besicovitch language pseudo-metric .................................. 50
  3.1 A Besicovitch pseudo-metric on language spaces ............................... 53
  3.2 The Besicovitch language norm is surjective ................................... 56
  3.3 Besicovitch distance quotient space ......................................... 58
    3.3.1 The Besicovitch distance equivalence relation and induced quotient space . 59
    3.3.2 The metric quotient topology .................................. 61
    3.3.3 Convergence in the quotient space ................................ 64
List of Tables

1. Language distances between small languages .......................... 126
2. Language distances between large languages .......................... 127
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Isometry between Besicovitch language space and quotient space</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>The Besicovitch quotient spaces.</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>The relationship of the Besicovitch language space, quotient space, and unit interval (upper quotient space).</td>
<td>77</td>
</tr>
<tr>
<td>4</td>
<td>Word in a generally testable language</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>The quotient spaces of the entropic distance.</td>
<td>106</td>
</tr>
</tbody>
</table>
Topology, Morphisms, and Randomness
in the Space of Formal Languages

David E. Kephart

ABSTRACT

This paper outlines and implements a systematic approach to the establishment, investigation, and testing of distances and topologies on language spaces. The collection of all languages over a given number of symbols forms a semiring, appropriately termed a language space. Families of languages are defined by interrelations among words. The traditional classification begins with the syntax rules or grammar of the language, that is, the word-transformations by which the entire language can be produced from a single axiom, or starting word. The study of distances between languages as objects and of the topologies induced by language distances upon spaces of languages has been of a limited character. Known language distances introduce topologically awkward features into a language space, such as total disconnectedness. This dissertation examines the topologies induced by three language distances, the effect that each one has upon the notion of a random language, and discusses continuity and word-distribution of structure-preserving language transformations, i.e., morphisms.

This approach starts from metric-like requirements, but adduces an additional condition intuitively appropriate to gauging language distance. At the same time, strict, i.e. non-metric pseudo-metrics are admitted as possible language distance functions, and these are investigated by the use of metric quotient spaces. The study of the notion of randomness implied by the topology induced by such a pseudo-metric on a language space offers insight into the structure of language spaces and verifies the viability of the pseudo-metric.

Three language pseudo-metrics are studied in this dissertation: a version of the most commonly-used (Cantor) word metric; an upper-density (Besicovitch) pseudo-metric borrowed from the study
of cellular automata; and an adaptation and normalization of topological entropy, each evaluated on the symmetric set-difference between languages. It is shown that each of these distances induces a distinct topology on the space of languages. The topology induced by Cantor distance is compact and totally disconnected, the topologies induced by the other two are non-compact, with entropic distance resulting in a topology that is the strict refinement of the Besicovitch topology, enhancing the picture of the smaller languages in the Besicovitch topology. It is also shown that none of the three topologies gives quantitative expression to the distinction between regular and linear languages, although, using Martin-Löf randomness tests, it is shown that each pseudo-metric is associated with a new notion of a random language.

A classification of language mappings is introduced, with the aim of identifying those which best preserve the structure of languages under specific topologies. There are results regarding continuity of mappings, the matrix representation of the pre-image of certain morphisms, and the formal expressions of the probability distribution of the image of certain morphism. The continuity of an injective morphism on its image is demonstrated under limited conditions.

Finally, the questions which this approach leaves open are detailed. While basic facts about a permutation-invariant version of symmetric set difference are shown, this has yet to be fully elaborated. The outline is presented for a metric which distinguishes between regular and linear languages by brute force. Syntactic and as algebraic topological continuations of this approach await investigation. A variation of the Cantor distance is introduced, and this induces a non-Cantor topology on a language space.

In summary, this dissertation demonstrates that it is possible to systematically topologize the formal language space, and, having done so, to determine the major effects this has upon the notion of random languages and upon language morphisms.
Chapter 1

Formalizing distance and topology on a space of languages

1.1 Are languages the support of a meaningful distance?

The investigation which resulted in this dissertation began three years ago. It derives from a seemingly simple question: how far apart are two languages? This emerged as a practical issue in the course of work on what is known as the word problem in DNA computing. Since DNA computation involves encoding data in oriented strands of nucleotides, there is a problem of determining which collections of words over the alphabet of nucleotides, \( \{A, C, G, T\} \), best satisfy the needs of computation. Those collections will “best satisfy” no concatenation of elements of which, other than for a very limited number of nucleotides, will bind to any other concatenation. The set of all possible concatenation of a collection of words must, first of all, form a free monoid over those words. Then what we are demanding, for the solution of the word problem, is that the free monoid over the word collection we have chosen be as distinct as possible, as far as possible, from the Watson-Crick complement of that free monoid. If this condition is not met, then, in the course of computation, while all DNA is single stranded, the familiar double helix bonds will form between significant portions of words encoding data, and that data will be lost for purposes of future computation. The question at the center of this seemed to be: what is the distance between a DNA language and its complement? In the event, this problem was tackled algebraically, modelling complementation as an antimorphic involution on words, as can be seen in [20] and [22]. The issues of what distance, what metric, and, therefore, what topology is appropriate to languages went unanswered.

The proper domain of the study of the distance between languages is, of course, the theory of formal languages. Perhaps best recognized for its integral role in theoretical computer science, formal language theory also lends results to scientists in a spectrum of disciplines, such as biomolecular computing, bioinformatics, and physics. Wherever it is appropriate to encapsulate information in non-commutative sequences, formal languages may be of use. Formal language theory originates
from the investigation of patterns in sequences, from the work on mathematical foundations and logic in the first half of the twentieth-century, from the coeval theory of computing and, in its specific modern form, from the classification of grammars worked out by Noam Chomsky [35] and Schützenberger in the 1960s. The initial aim, from these several standpoints, was to classify and characterize languages as mathematical objects. Each language is one of uncountably many members of a countable collection of semirings (see, for instance, [24]). A language belonging to one of a classical hierarchy of families is recognized by a grammar, i.e., syntax, or a logical machine of a particular type. Simpler syntactic relations typify sub-families of languages. Still other collections of languages are identified by algebraic structure. Operations, such as morphisms of words (as in [14]) or splicing rules, characterize still other families of languages. Notable for its lack of extensive investigation is the question of the topology of languages, the distance between languages, and whether and how a new classification of languages can be achieved or previous classifications synthesized by correctly metrizing the language space. Without a systematic study of language topology there is no reason to accept the claims of any particular language distance offered for consideration. Yet, taking the DNA word problem as an example, where vast numbers of interactions in a system depend upon accurately gauging the distance-like relationship between its elements, it can be seen that topology is a vital issue in any application of language theory to complex systems.

This dissertation, then, is part of an attempt to fill a gap in language theory with a rigorous approach to evaluating language distance. We are hardly alone in recognizing the curious lack of a standard language distance. The compilation worldwide of vast databases, like the protein database maintained in the US, which is distinctly syntactic in character, has necessitated the development of ad hoc methods of evaluating language distances. These vary from the metric on infinite sequences used in symbolic dynamics, to the more fashionable-sounding Normalized Google Distance (NGD) which is shown in [10] to be the “best possible” for the amorphous database of the Web. Both of these are, however, word metrics, although the NGD relies on the statistical proximity of words in an immense language (the Google corpus), and is one particular application of the the similarity metric and discussed discussed in [8]. As another example, in [36], a measure is imposed on regular languages to quantify the degree of correspondence of the languages of plant automata to that of the supervisor. From the topological side, we found “circumstantial” evidence that our goal is realizable in such work as [15], where the authors successfully apply algebraic topology directly to problems in
computing theory. Finally, in [6], Carbone and Gromov say they aim to obtain “a possible formalism of combinatorial and numerical (entropic) structures on spaces of sequences which reflect, up to some degree, the organization and functions of DNA and proteins,” which amounts to a statement of our goal in the realm of (natural) DNA languages. We feel that the appearance of such distances as those listed above and the search for evidence of functional relationships in the quantifiable aspects of languages are two sides of the same issue, namely, that a more generalized theory of language topology is attainable.

In order to assemble machinery for the investigation of language distances and the language space topologies they induce we turned to resources, first of all, in symbolic dynamics. In Lind and Marcus [26] and the much-cited work of Kitchens [23], a language and the relations within it are viewed as the product of the inherent operation of a dynamical system. These works employ the metric on biinfinite sequences, which can be transformed into a language metric, namely, the one we will call here Cantor distance. Our primary source for details of the Cantor distance has been the work of Genova and Jonoska on forbidding and enforcing systems[13], but earlier sources for this natural metric exist. We cite the paper of Vianu (1977) [38], which in turn refers to a paper by Bodnarchük, and both discuss how to metrize a language space as a normed linear space. Our first alternative to the Cantor distance is an analog of the pseudo-metric of Cattaneo, Formenti, Margara, and Mazoyer [7]. The name “Besicovitch pseudo-metric” is, in fact, borrowed from the further discussion of this distance [3, 12]. This is not a language pseudo-metric at all, but, rather, a metric on the biinfinite sequences that make up the elements of the configuration space of cellular automata. There is some similarity of this pseudo-metric to the $\zeta$ function on languages, which is less useful than topological entropy in understanding language behavior, as discussed, for instance, in [23] (but well-known also in information theory). What we call the entropic distance arose from the consideration that using an aspect of languages known to give better results, might similarly give rise to a better language distance.

Armed with several examples of possible language pseudo-metrics, we next considered ways in which a language topology might be tested. The ability to distinguish metrically (rather than syntactically) the classical language families seemed a reasonable goal. If it is not attained, we needed, as well, a characteristic which likely to be available for analysis in any language topology. The validity of a language topology seemed to hinge, ultimately, on the following: does a given distance
and topology obscure or enhance the relative amount of information expressible in a language? In other words, we ask how the complexity of languages changes with a change in topology. We found a powerful resource for working with this in the papers of Calude, et al., [4, 5], Hertling and Weihrauch, [16], and in the seminal work on randomness tests by Martin-Löf [29] and Kolmogorov [25].

This dissertation, therefore, investigates a formalization of language distance. It establishes criteria for a particular type of distance, determines to what degree the topologies induced by three qualifying distances allow for the reflection of known distinctions between languages as theoretical objects, and in what respect these topologies give a new picture of language complexity. Finally, it offers some analysis of the behavior of language transformations under alternate topologies.

In this first chapter, we give the definitions upon which we base our investigation of language distances, including definitions of language spaces, language pseudo-metrics, language norms, language morphisms, and randomness tests, as well as the notation appropriate to their discussion. In addition, we make some basic observations about classes of language mappings: for example, Lemma 1.2 shows that a natural isomorphism between language spaces exists only when the spaces are in all essential respects identical. In the three ensuing chapters we consider the three language norms and pseudo-metrics mentioned above. In the fifth chapter, we discuss further aspects of language morphisms and language distribution under morphisms. In the last chapter, we draw some conclusions and outline of the work remaining to be done and the open questions regarding language topology.

The principle conclusion of this dissertation is that, by a systematic approach, we can create widely divergent language topologies which allow for ready theoretical examination, but that this is the threshold to a comprehensive study of the space of language pseudo-metrics and distances in general. We demonstrate the differences and difficulties in determining the way pseudo-metrics define randomness. We show how they locate certain language families, and how this points to a means of classifying languages and determining their complexity without prior knowledge of their syntax. The possibility that the dynamics of languages, too, is dependent upon the topology employed is demonstrated by showing certain relationships between topology and the behavior of language transformations. In this way, a link is established between topology, morphisms and
randomness on the space of formal languages and the comprehension of any process which may be modelled as the operation of a discrete time dynamical system.

1.2 Basic definitions

In this section, we state the definitions, notions and notation from formal language theory which will be employed in this dissertation. All notation introduced is adapted, and, in some cases, modified for the discussion of languages spaces, morphisms, pseudo-metrics, and randomness.

In subsection one the basic objects from general mathematics, set theory, and language theory are listed. The second subsection defines languages, language operations, and the classical hierarchy of syntax-grammars of languages as well as several other families of languages which will be used later. The third subsection defines notions specific to language spaces.

1.2.1 General notation

The primary object from which language theory begins is an alphabet. Alphabets, denoted $A, B, C,...$, are finite non-empty sets. The elements of an alphabet $A$ are called its symbols. Unless otherwise noted, we will assume we are discussing alphabet $A$.

The symbols $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ denote the nonnegative integers, the integers, and the real numbers. For convenience, in place of the expression “$1 \leq i \leq k$” we will say “$i \in \mathbb{N}_k$”, so that $\mathbb{N}_k$ will be used occasionally to denote the set $\{1, 2, 3, \ldots, k\}$ for $k \in \mathbb{N}$. Then $\mathbb{N}_0$ is just another way of denoting $\emptyset$.

The cardinality of the set $S$ is denoted $\#S$, and the collection of the subsets of $S$, its power set, is denoted $\mathcal{P}(S)$. Given sets $S$ and $T$, we denote by $S \triangle T$ the symmetric set difference of $S$ and $T$, namely, the set $S \triangle T = S \setminus T \cup T \setminus S = (S \cup T) \setminus (S \cap T)$. If $f$ is a function mapping set $S$ to set $T$, we will write $f : S \to T$. If the $f$ is injective, we will write $f : S \hookrightarrow T$; if $f$ is surjective, we will write $f : S \twoheadrightarrow T$; and, if $f$ is a bijection, we will write $f : S \leftrightarrow T$.

Unless otherwise indicated, we will assume that $\#A$, the cardinality of the alphabet under discussion is $\alpha$, and that $\alpha > 1$.

A word $w$ over the alphabet $A$ is a finite sequence of symbols of $A$, a mapping $w$ from $\mathbb{N}_k$ into $A$ for some $k \in \mathbb{N}$. Then the length of word $w$ is $k$ and is denoted $|w|$. The Parikh count of symbol $a \in A$ in the word $w$, the number of occurrences of symbol $a$ in $w$, is denoted $|w|_a$. Subscripts in
square brackets represent the excerption of symbols from a word. That is, \( w[i] \) will denote \( w(i) \), the \( i \)th symbol of \( w \), while \( w[i, j] \) will denote the word formed of symbols \( i \) through \( j \) of word \( w \). The empty word will be denoted \( \lambda \).

Then any two words \( u \) and \( v \) can be concatenated, forming the word \( uv \), where \( |uv| = |u| + |v| \). If the concatenation \( uvw = x \), where \( u \), \( v \), and \( w \) are (possibly empty) words over \( A \), the word \( u \) is a prefix of \( x \), \( v \) is a factor of \( x \), and \( w \) is a suffix of \( x \). The set of all factors of word \( w \) is denoted \( \text{Fac}(w) \), the set of all prefixes of \( w \) is denoted \( \text{Pref}(w) \), and the set of all suffixes of \( w \) is denoted \( \text{Suf}(w) \).

The set of all words over alphabet \( A \) is denoted \( A^* \), whereas the set of all words of positive length is denoted \( A^+ = A^* \setminus \{\lambda\} \). Thus the sets \( A^* \) and \( A^+ \) are the free monoid and free semigroup, respectively, generated by \( A \) under the operation of concatenation. If word \( w \in A^* \), then \( w^0 = \lambda \), and \( w^k = w w^{k-1} \), for \( k > 0 \).

A function mapping symbols to words will be called a literal mapping, and a function mapping words to words will be called a word mapping. Every literal mapping extends to a word mapping which is a word morphism, under which, for all \( u, v \in A^* \), the mappings \( \varphi(uv) \) and \( \varphi(u) \varphi(v) \) are equal, and where \( \varphi(\lambda) = \lambda \). If, for word morphism \( \varphi : A^* \rightarrow B^* \), \( \varphi(w) = \lambda \) if and only if \( w = \lambda \), then \( \varphi \) is called a nonerasing word morphism. A code is usually defined as a word morphism which is uniquely factorable over the alphabet. That is, if \( \varphi \) is a code, and

\[
\varphi(w_1) \varphi(w_2) \varphi(w_3) \cdots \varphi(w_r) = \varphi(v_1) \varphi(v_2) \varphi(v_3) \cdots \varphi(v_s),
\]

for \( w_1, w_2, \ldots, w_r, v_1, v_2, \ldots, v_s \in A \), then \( r = s \) and \( w_i = v_i \) for \( i \in \mathbb{N}_r \). But then, without ambiguity, a code is an injective word morphism.

Our interest is in collections of words, languages, which we define next.

1.2.2 Languages and language families

A major achievement in language theory was the identification of a hierarchy of nested families of languages using syntax grammars, accomplished by Chomsky in the 1960s. Following the definition of basic operations on languages, we recall here the definitions of the classical language families. Other language classification systems such as L-systems and splicing systems will not be dealt with in this dissertation.
Any subset of $A^*$ is called a language over $A$, and languages will be specified with upright characters: $L$, $M$, $N$, . . . The symbol $\Lambda$ will denote the empty set understood as the empty language. A language which does not contain the empty word $\lambda$ is called $\lambda$-free. Languages can be concatenated: for languages $L$ and $M$, the concatenation of $L$ and $M$, denoted $LM$, is the language

$$LM = \{ uv \in A^* : u \in L, v \in M \}.$$  

The complement of language $L$, denoted $L^c$, is the language $A^* \setminus L$. Collections of languages will be denoted in boldface, $L$, $M$, $N$, . . . For any language $L$, $L^0 = \{\lambda\}$ and $L^k = LL^{k-1}$ for all $k > 0$. The Kleene-+ and Kleene-\+-operations on words, which extend to operations on languages in the same manner as does concatenation, generate languages: the symbol $w^*$ denotes the language $\{w^k : k \in \mathbb{N}\}$ and the symbol $w^+$ denotes the language $w^* \setminus \{\lambda\}$, where $w \in A^*$. Likewise, the Kleene-* operation on language $L$ generates the language $L^* = \{L^k : k \in \mathbb{N}\}$ and the Kleene-+ operation on $L$ generates the language $L^+ = L^* \setminus \{\lambda\}$. For $k > 0$, by $L^{<k}$ we denote the union of $L^0$, $L$, $L^2$, . . . up to $L^{k-1}$, for $k > 0$. In particular, $\{#A^k\}_{k \in \mathbb{N}} = \{\alpha^k\}_{k \in \mathbb{N}}$, so that

$$#A^{<k} = \frac{\alpha^k - 1}{\alpha - 1}$$  

The classical language families are defined in terms of grammars that generate, or recognize them. A grammar $G$ is a 4-tuple, $G (N, T, \delta, S)$, where $N$, the non-terminal symbols, and $T$, the terminal symbols, are disjoint alphabets used to produce a language, and $S \in N$ is an axiom which initiates the production of the language. The rules of production are given by the finite relation $\delta$, which is a subset of $(N \cup T)^* N (N \cup T)^* \times (N \cup T)^*$ such that $\delta \cap [(N \cup T)^* \times T^*] \neq \emptyset$. If the word pair $(u, v) \in \delta$ we write $u \rightarrow v$. If $x, y \in (N \cup T)^*$ and $x = x_1ux_2$, $y = y_1vy_2$ such that $u \rightarrow v$, we write $x \Rightarrow y$. Denote by the symbol $\Rightarrow^*$ the reflexive and transitive closure of $\Rightarrow$. Then the language recognized by the grammar $G$, denoted $L(G)$, is just the set $L(G) = \{u \in T^* : S \Rightarrow^* u\}$.

The family of all languages recognized by grammars is the family of Type 0 languages ($\mathcal{L}_0$). It is a standard proof that the Type 0 languages are in fact the recursively enumerable languages, viz., those languages recognized by Turing Machines (See, e.g., pp. 178-9 in [35]). We will use the designation RE for this family. This establishes an intimate relationship between language classification and theoretical computer science.
If a grammar \( G(N,T,\delta,S) \) is such that, if \( u \rightarrow v \), then \(|u| \leq |v|\), except that it may be that \( S \Rightarrow \lambda \), but only if \( \delta \subset ((N \cup T)^* \times ((N \cup T)\setminus S)^*) \), then \( G \) recognizes a context-sensitive language\(^1\), and \( L(G) \) belongs to the family of Type 1 languages \((\mathcal{L}_1)\). This family we will denote \( \text{CS} \).

If a grammar \( G(N,T,\delta,S) \) is such that, if \( u \rightarrow v \), then \( u \in N \), then \( G \) defines a context-free grammar. Then \( L(G) \) is belongs to the family of Type 2 languages \((\mathcal{L}_2)\). We will denote this family \( \text{CF} \).

If a grammar \( G(N,T,\delta,S) \) is such that, if \( u \rightarrow v \), then \( u \in N \) and \( v \in T^* \cup (T^*NT^*) \), then \( G \) defines a linear grammar, and \( L(G) \) belongs to the family of linear languages, denoted \( \text{LIN} \).

If \( G(N,T,\delta,S) \) is such that \( u \rightarrow v \) only if \( u \in N \) and \( v \in T \cup TN \cup \{\lambda\} \), then \( G \) defines a regular or rational grammar and \( L(G) \) is a member of the family of Type 3 languages \((\mathcal{L}_3)\), which we will denote \( \text{REG} \).

It is known that \( \mathcal{L}_0 \supseteq \mathcal{L}_1 \supseteq \mathcal{L}_2 \supseteq \text{LIN} \supseteq \mathcal{L}_3 \). There are alternative characterizations of all of these families. The Myhill-Nerode Theorem and Pumping Lemma, for instance, are well-known means of identifying regular languages, and will be referred to in this paper without further detail.

There are two prominent subfamilies of \( \text{REG} \) (the regular languages) of interest to us. The first is the family of locally testable languages, denoted \( \text{LOC} \). If a language \( L \) is in \( \text{LOC} \), then there is a fixed integer \( k \in \mathbb{N} \), called a window length, and a proper subset \( F \) of \( A^k \) such that, if every factor of \( w \) of length \( k \) is in \( F \), then \( w \) belongs to \( L \). Thus, \( L \) is a locally testable language if “\( w \in L \)” is decidable merely by inspecting each \( k \)-length factor of \( w \). As an example, suppose \( A = \{0,1\} \). Then the language \( L_1 = \{\lambda,0\}\ (10)^*\ \{1,\lambda\} \) is a local language, but not the language \( L_2 = (10)^* \).

For, although every two-symbol factor of every word in \( L_2 \) is an element of the set \( \{10,01\} \subseteq A^2 \), the relation \( \text{Fac}(w) = \{10,01\} \) does not guarantee that \( w \in L_2 \), since the (non-local) condition \( w[1] = 1 \) and \( w[|w|] = 0 \) is also necessary. The second subfamily of importance is the family of finite languages, denoted \( \text{FIN} \). For each language \( F \) in \( \text{FIN} \), there is a maximum integer \( n \) such that \( F \cap A^n \neq \Lambda \).

---

\(^{1}\)This is actually the definition of a monotonous grammar. In a context-sensitive grammar, \( u \rightarrow v \) if there is at least one non-terminal symbol in \( u \) and \( v \neq \lambda \) (except if \( S \rightarrow \lambda \) and \( S \) is not a factor of \( v \) for any \( u \rightarrow v \in P \)) but context-sensitive and monotonous grammars recognize the same languages.
If it is necessary to restrict the consideration of one of these families to those members of the family which are languages in a given language space, we will use a subscript to denote the alphabet involved. For example, the set of all regular languages over the alphabet $A$ form the family $\text{REG}_A$.

Languages may additionally be distinguished as factorial, prolongable, or transitive. A language $L$ is factorial if all of the factors of each word $w$ in $L$ are also words in $L$, that is, $w \in L$ implies $\text{Fac}(w) \subseteq L$. A language is prolongable if, for each word $v \in L$, there exist words $u, w \in A^+$ such that the words $uv$ and $vw$ are in $L$. Finally, a language $L$ is transitive if, for each pair of words $u, w \in L$ there is a word $v \in A^*$ such that the word $uvw \in L$. Obviously, any transitive language other than $\{\lambda\}$ is prolongable.

Finally, we note the following special languages formed by left, right, and two-sided monoid ideals of $A^*$.

1. A language $I \in \mathcal{P}_A$ is called a right ideal of $A^*$ if $IA^* \subseteq I$.
2. A language $I \in \mathcal{P}_A$ is called a left ideal of $A^*$ if $A^*I \subseteq I$.
3. A language $I \in \mathcal{P}_A$ is called a two-sided ideal of $A^*$ if $A^*IA^* \subseteq I$.

Where the notion of the size of a language is not otherwise defined, the size of a language will mean its cardinality. Thus, the size of a language is a natural number if the language is finite, and is infinite (actually, $\omega$), otherwise.

1.2.3 Language spaces

The collection of all possible languages is a set, and we call this collection the formal language space. The collection of all languages over a given alphabet $A$ is then the language space over $A$, the collection of all subsets of $A^*$, i.e., the power set $\mathcal{P}(A^*)$. We will denote the language space over $A$ by $\mathcal{P}_A$. As a power set, a language space inherits the set operations on languages over $A$. In particular, $\mathcal{P}_A$ is a commutative monoid under the set operation of union.

Denoting by $\cdot$ the operation of language concatenation, the triple $(\mathcal{P}_A, \cdot, \{\lambda\})$ is a monoid. Since language concatenation is distributive over set union, the 5-tuple $(\mathcal{P}_A, \cup, \cdot, A, \{\lambda\})$ is a semiring. Moreover, a language space $\mathcal{P}_A$ is closed under countable unions and set differences, so that the language space over $A$ is a $\sigma$-ring on the set $A^*$. Thus the unusual situation exists that, if a measure is defined on $A^*$, then every element of the language space $\mathcal{P}_A$ becomes a measurable set.
In the section on morphisms, it will develop (see Theorem 1.2 on page 21) that language spaces over alphabets with the same number of symbols are indistinguishable. We will occasionally make use, therefore, of the canonical language space over an alphabet with \( k \) symbols, where \( k \in \mathbb{N} \), which is just the power set of \( k \) natural numbers, \( \mathcal{P}(\mathbb{N}^*_k) \). The canonical language space over \( k \) symbols will be denoted \( \mathcal{P}_k \).

In addition, to conveniently refer to sections of a language we will use the following notation, where we understand a language as being sectioned into words of different lengths.

**Definition 1.2.1** If \( L, M \in \mathcal{P}_A \) and \( k \in \mathbb{N} \), then

1. let \( L[k], L[<k], \) and \( L[\leq k] \) denote the sets \( L \cap A^k, \bigcup_{i=0}^{k-1} L[i] \), and \( \bigcup_{i=0}^{k} L[k] \), words in language \( L \) of exactly, up to, and up to and including length \( k \), where \( k \in \mathbb{N} \);

2. let \( \#L[k], \#L[<k], \) and \( \#L[\leq k] \) denote the cardinalities \( \#(L \cap A^k), \#\bigcup_{i=0}^{k-1} L[i] \), and \( \#\bigcup_{i=0}^{k} L[k] \) of the above sets; and

3. let \( L \triangle^k M, L \triangle^k M, \) and \( L \triangle^k M \) denote the sets \( (L \triangle M) \cap A^k = (L \triangle M)[k], (L \triangle M)[<k], \) and \( (L \triangle M)[\leq k] \), i.e., sections of the symmetric set-difference of the languages \( L \) and \( M \).

There is a qualitative change in passing from the countable monoid \( A^* \) to the uncountable collection of all languages over a given alphabet. In fact, the monoid \( (\mathcal{P}_A, \{\lambda\}, \cdot) \) is not finitely generated. Even the question of the commutativity of languages poses serious, possibly intractable problems. In the discussion in [11], for example, a fixed point characterization of the centralizer of a language is given which, however, may not finish in finitely many iterations, even for finite languages. To restrict our discussion to manageable mappings between languages, the following section isolates transformations between language spaces which will always preserve major aspects of language structure.

### 1.3 Morphisms on words and language spaces

We will need to discuss issues like compactness, continuity, and dynamics, all of which involve structure-preserving mappings between spaces. The first subsection defines mappings and morphisms on languages. The term *language morphism* will be reserved for a semiring morphisms from
one language space into another. The distance between two languages should have a relationship
to the distance between the images of the two languages. It will develop that it is of concern to, first,
prescribe conditions on morphisms which prevent the deletion of distinctions between languages,
and, second, prescribe conditions which prevent the propagation of language differences. Toward
the first goal, the notion of a nonerasing morphism is extended from word to language mappings,
and, toward addressing the second goal, the complementary notion of a nonexpansive mapping is
introduced. For languages, both terms refer to cardinality of languages rather than, as in the case of
words, to word-length. In particular, the second concept would be of little use with regard to word
morphisms. The subsection also settles the issues of injectivity and of the conditions under which a
morphism is the extension of a word morphism.

In the second subsection, we define the notion of an isomorphism between language spaces, a
natural isomorphism of language spaces, and automorphisms of language spaces. We prove that a
natural isomorphism exists only between language spaces that are, for practical purposes, the same.

The third subsection introduces a notion of the symmetric set difference minimized over the auto-
morphic images of two languages. Our language pseudo-metrics will involve the use of symmetric
set differences. The most common objection to the use of the symmetric set-difference is that, under
such a distance, simply by permuting the symbols in the alphabet, a language may be derived with
the identical syntactic features as the original, but at perhaps maximum distance from the original.
With a permutation-invariant set-difference, this objection is overcome.

1.3.1 Extending literal mappings and word morphisms to language morphisms

Where possible, we extend notions commonly applied to word mappings to notions describing map-
pings between language spaces. A language mapping $\Phi$ is a set function mapping language space
$P_A$ into language space $P_B$, where possibly $A = B$. The characteristics of most language map-
pings are hard to describe. For example, a given language mapping may map a finite language to a
larger or smaller language, or even to an infinite language, and may map infinite languages to finite
languages. We specify what restrictions must be made in order that the effects of a language map-
ping may be completely understood. We ask, first, that language mappings preserve concatenation,
and, second, that they preserve set union. We reserve the term language morphism for mappings
with these two properties, the semiring morphisms between language spaces. But this does not
guarantee that a mapping behaves “properly”. Even a semiring morphism need not map the identity \{\lambda\} of \mathcal{P}_A to \{\lambda\} in \mathcal{P}_B. To obtain a finitely generated and injective mapping, we require the notions of a nonerasing and nonexpansive mapping.

**Definition 1.3.1** A monoid morphism \( \Phi : \mathcal{P}_A \rightarrow \mathcal{P}_B \) is a language mapping from \( \mathcal{P}_A \) to \( \mathcal{P}_B \) which preserves concatenation. That is, \( \Phi(L)\Phi(M) = \Phi(LM) \) for all \( L, M \in \mathcal{P}_A \). A language morphism is a monoid morphism \( \Phi : \mathcal{P}_A \rightarrow \mathcal{P}_B \) which is a semiring morphism on \((\mathcal{P}_A, \cdot, \cup, \lambda, \Lambda)\), i.e., if \( \Phi \) preserves both the concatenation and the finite unions of languages. If \( \Phi \) is a language morphism, and \( L \) is a collection of languages, then \( \Phi(\{w : w \in L, \text{ for some } L \in L\}) = \{v : v \in \Phi(L), \text{ for some } L \in L\} \).

A language mapping \( \Phi : \mathcal{P}_A \rightarrow \mathcal{P}_B \) will be called

- **nonerasing** if there is a bijection between \( L \) and a subset of \( \Phi(L) \) for every language \( L \in \mathcal{P}_A \);
- **nonexpansive** if, for every finite language \( F \), \( \#\Phi(F) \leq \#F \); and
- **trivial** if there is a language \( M \in \mathcal{P}_B \) such that \( \Phi \) maps all languages in \( \mathcal{P}_A \) to \( M \).

An injective monoid morphism will be called a language space code; an injective language morphism will be called a language code.

Note that if language mapping \( \Phi \) is both nonerasing and nonexpansive, and if \( F \) is a finite language, then \( \#\Phi(F) = \#F \). The distinction between a monoid morphism, which preserves concatenations of languages but not necessarily the unions of languages, and a language morphism, which preserves both, is illustrated by the following fact.

**Fact 1** For any language morphism \( \Phi : \mathcal{P}_A \rightarrow \mathcal{P}_B \) and any non-empty language \( M \in \mathcal{P}_A \),

\[
\Phi(M) = \bigcup_{w \in M} \Phi(\{w\}).
\] (1.2)

This means that countable union is preserved. Moreover, if language \( N' \) is a non-empty finite subset of \( \Phi(M) \), then there is a finite language \( N \subseteq M \), such that \( \Phi(N) \supseteq N' \).

Note that it is not true that, however, that, for any finite language \( N' \subseteq \Phi(N) \), there exists (finite or infinite) language \( N \) such that \( \Phi(N) = N' \). For example, suppose \( A = \{a, b, c\} \) and that \( \Phi : \mathcal{P}_A \rightarrow \ldots \)
A language morphism $\Phi$ is such that

\[
\begin{align*}
\Phi : \{a\} & \mapsto \{a\} \\
\Phi : \{b\} & \mapsto \{b, c\} \\
\Phi : \{c\} & \mapsto \{ab, ba\}.
\end{align*}
\]

Then $\Phi (\{ab\}) = \Phi (\{a\}) \Phi (\{b\}) = \{ab, ac\}$. It can be seen that, although $\Phi (\{ab, c\}) = \{ab, ac, ba\}$ and the language $\{ac, ba\}$ is a subset of $\Phi (\{ab, c\})$, there is no subset of $\{ab, c\}$, indeed, there is no subset $N$ of $\{a, b, c\}^*$ such that $\Phi (N) = \{ac, ba\}$.

The behavior of a nonerasing language morphism is characterized by its behavior on the finite languages.

**Fact 2** A language morphism $\Phi$ is nonerasing if, for every finite language $F$, $\#\Phi (F) \geq \#F$.

**Proof.** Suppose, toward contradiction, that the language morphism $\Phi : \mathcal{P}A \to \mathcal{P}B$ meets the hypothesized conditions, but is not nonerasing because it maps some infinite language $L$ to a finite language $G$. Then certainly there is a finite subset $F$ of $L$ such that $\#F > \#G$. By Fact 1, $\Phi (F) \subseteq G$, and so $\#F > \#\Phi (F)$, contrary to hypothesis. $\square$

Notice that it is possible that, under a monoid morphism, the image of the empty language is non-empty and the image of $\{\lambda\}$ is not $\{\lambda\}$. The following is an example of this.

**Example 1** As an example, define the language mapping $\tilde{\Phi} : \mathcal{P}2 \to \mathcal{P}2$ such $\tilde{\Phi}$ takes $\Lambda$ to $\mathbb{N}_2^*$ and languages other than $\Lambda$ to the language consisting of all words containing at most as many instances of the symbol $2$ as the length of the shortest word in the language. That is, let $\tilde{\Phi} (\{\lambda\}) = \{1, 2\}^*$, $\tilde{\Phi} (\{\lambda\}) = 1^*$, and for non-empty $L \subseteq \{1, 2\}^+$, $\tilde{\Phi} (L) = \{w \in \mathcal{P}2 : \|w\|_2 \leq m_L\}$, where $m_L$ is the length of the shortest word in $L$. Then, for $L, M \subseteq \{1, 2\}^+$,

\[
\tilde{\Phi} (\{\lambda\}) \tilde{\Phi} (L) \tilde{\Phi} (\{\lambda\}) \tilde{\Phi} (M) \tilde{\Phi} (\{\lambda\}) = \{w \in \mathcal{P}2 : \|w\|_2 \leq m_L + m_M\}
\]

and, since $\lambda \in \tilde{\Phi} (L)$ for all $L \subseteq \{1, 2\}^+$, we also have that $\tilde{\Phi} (\Lambda) \tilde{\Phi} (L) = \tilde{\Phi} (L) \tilde{\Phi} (\Lambda) = \tilde{\Phi} (\Lambda) = \{1, 2\}^*$. Then $\tilde{\Phi}$ is a monoid morphism from $\mathcal{P}2$ to $\mathcal{P}2$. 

13
In certain trivial cases, it is obvious that monoid and language morphisms will produce such results. For instance, if \( \Phi \) which maps every language in \( \mathcal{P}_A \) to a language \( M \in \mathcal{P}_B \) such that \( MM = M \) is a monoid morphism which, if \( M \neq \Lambda \) and \( M \neq \{\lambda\} \), neither maps \( \Lambda \) to \( \Lambda \) nor \( \{\lambda\} \) to \( \{\lambda\} \). As the following example shows, even a language morphism need not map \( \{\lambda\} \) to \( \{\lambda\} \).

**Example 2** If we define the language mapping \( \tilde{\Phi}' \) to map \( \Lambda \) to \( \Lambda \), so that

\[
\tilde{\Phi}'(L) = \begin{cases} 
\Lambda, & \text{if } L = \Lambda, \text{ and} \\
\tilde{\Phi}(L), & \text{otherwise,}
\end{cases}
\]

where \( \tilde{\Phi} \) is the monoid morphism from Example 1, then

\[
\tilde{\Phi}'(L \cup M) = \{w \in \mathcal{P}_A : |w|_2 \geq \min \{m_L, m_M\}\} = \tilde{\Phi}'(L) \cup \tilde{\Phi}'(M)
\]

for all \( L, M \in \mathcal{P}_2 \), and hence \( \tilde{\Phi}' \) is a language morphism.

The following fact lists the restrictions which result in better behavior on the part of monoid and language morphisms.

**Fact 3** Every non-trivial monoid morphism \( \Phi : \mathcal{P}_A \to \mathcal{P}_B \) maps \( \Lambda \) to either \( \Lambda \) or to an infinite language and maps \( \{\lambda\} \) to either \( \{\lambda\} \) or an infinite language. If \( \Phi \) maps any language in \( \mathcal{P}_A \) to a finite language in \( \mathcal{P}_B \), then \( \Phi(\{\lambda\}) = \{\lambda\} \). A language morphism \( \Phi : \mathcal{P}_A \to \mathcal{P}_B \) which maps any language in \( \mathcal{P}_A \) to a finite language in \( \mathcal{P}_B \) maps \( \Lambda \) to \( \Lambda \).

**Proof.** If \( \Phi \) is a monoid morphism, then \( \Phi(\Lambda) \Phi(\Lambda) = \Phi(\Lambda) \), so \( \Phi(\Lambda) \) can be no finite, non-empty language other than \( \{\lambda\} \). But, if \( \Phi(\Lambda) = \{\lambda\} \), and \( L \in \mathcal{P}_A \), then we have

\[
\Phi(\Lambda) = \Phi(L\Lambda) = \Phi(L) \Phi(\Lambda) = \Phi(L) \{\lambda\} = \Phi(L),
\]

which implies that \( \Phi \) is not non-trivial, contrary to hypothesis. In addition, \( \Phi(\{\lambda\}) \Phi(\{\lambda\}) = \Phi(\{\lambda\}) \), so that \( \Phi(\{\lambda\}) \) can be no non-empty finite language other than \( \{\lambda\} \). If \( \Phi(\{\lambda\}) = \Lambda \), and \( L \in \mathcal{P}_A \), then we would have

\[
\Phi(L) = \Phi(L\{\lambda\}) = \Phi(L) \Lambda = \Lambda,
\]
and $\Phi$ is not nontrivial, contrary to hypothesis. Finally, if $\Phi$ is a language morphism and there exists a language $L \in \mathcal{P}_A$ such that $\Phi (L) = M$, where $M$ is a finite language in $\mathcal{P}_B$, then we have

$$M = \Phi (L) = \Phi (L \cup \Lambda) = \Phi (L) \cup \Phi (\Lambda) = M \cup \Phi (\Lambda),$$

implying that $\Phi (\Lambda) \subseteq M$. Thus, $\Phi (\Lambda)$ is finite. We have just shown that, because $\Phi$ is a monoid morphism, this means that $\Phi (\Lambda) = \Lambda$. \hfill $\square$

The other properties defined in Definition 1.3.1 each have a significance which can be first approximated by considering its negation. A language mapping is not nonexpansive if it maps a finite language $F$ either to an infinite language or to a finite language containing more words than $F$ does. A language mapping is not nonerasing if it maps an infinite language to a finite one, or a finite language to a finite language consisting of fewer words. These give sufficient control over language mappings to make some observations about injectivity, i.e., about language space codes and language codes.

First, we need to know how to extend a word morphism to a language morphism.

Fact 4 Every word morphism $\varphi : A^* \to B^*$ extends to the unique language mapping $\Phi_\varphi : \mathcal{P}_A \to \mathcal{P}_B$ such that

$$\Phi_\varphi (L) = \{ \varphi (v) : v \in L \} = \bigcup_{v \in L} \Phi (v).$$

(1.3)

and $\Phi_\varphi (\Lambda) = \Lambda$.

A construction which is the converse of Fact 4, i.e., for a given language morphism to find the word morphism of which it is the extension, is possible for nonexpansive language codes. We prove this after showing that language codes must be nonerasing, and that nonerasing and nonexpansive language morphisms are language codes.

Lemma 1.1 If a language morphism $\Phi : \mathcal{P}_A \to \mathcal{P}_B$ is nonerasing and nonexpansive, then it is a language code. Every language code is nonerasing.

Proof. Let $\Phi : \mathcal{P}_A \to \mathcal{P}_B$ be a non-trivial, nonerasing, nonexpansive language morphism. Suppose, toward contradiction, that $\Phi$ is not injective, i.e., not a language code. Then there exist languages $L, M \in \mathcal{P}_A$ such that $\Phi (L) = \Phi (M)$ and $L \neq M$. Since $\Phi$ is nonerasing, there exists a
subset $F$ of $L$ and a subset $G$ of $M$, $F \neq G$, such that $F$ and $G$ are distinct finite languages which have the same image under $\Phi$. Then $\Phi (F) = \Phi (G) = H$, where $H$ is some finite language in $\mathcal{P}_B$, since $\Phi$ is nonexpansive. Therefore, there exists $k \in \mathbb{N}$ such that $#F = #G = #H = k$, since $\Phi$ is both nonexpansive and nonerasing. Then $\Phi (F \setminus (F \cap G)) = \Phi (G \setminus (F \cap G)) = H' \in \mathcal{P}_B$ and, for some $k' \in \mathbb{N}$ such that $0 < k' \leq k$, we have that $#(\Phi (F \setminus (F \cap G))) = #(\Phi (G \setminus (F \cap G))) = #H' = k'$. But the cardinality of the union of the sets $F \setminus (F \cap G)$ and $G \setminus (F \cap G)$ is $2k'$, since they are disjoint, whereas $#(\Phi ((F \setminus (F \cap G)) \cup (G \setminus (F \cap G)))) = k'$, since $\Phi$ is nonexpansive and nonerasing. This is a contradiction. The conclusion is that $\Phi$ is a language code. This establishes the first claim.

Now suppose that a language code $\Phi : \mathcal{P}_A \to \mathcal{P}_B$ is not nonerasing. By Fact 2 we can assume that there exists $F \in \text{FIN}_A$ such that $#\Phi (F) < #F$. Since $\Phi$ is a language code, the cardinality of the set $F = \{\Phi (\{w\}) : w \in F\}$ is exactly $#F$. We claim, however, that there is a proper subset $F'$ of $F$ such that $\Phi (F') = \Phi (F)$. For by (1.2), $\bigcup F = \Phi (F)$, and, if $v \in \Phi (F)$, then there exists $w \in F$ such that $v \in \Phi (\{w\})$. However, by assumption, $#\Phi (F) < #F$. If $#\Phi (F) = k$, there is a collection $F'$ of at most $k$ elements of $F$ such that $\bigcup F' = \Phi (F)$. Thus, if $F' = \{w \in \mathcal{P}_A : \Phi (\{w\}) \in F\}$, then $F' \subsetneq F$ and $\Phi (F') = \Phi (F)$, as claimed.

But then $\Phi$ is not a code, because it is not injective, contrary to assumption.

It follows that $\Phi$ is nonerasing, which establishes the second claim of the lemma. □

Next, two possible converses of Lemma 1.1.

**Corollary 1.1.1** A language mapping that is nonerasing and nonexpansive and not a language code is not a language morphism.

An example of the type of badly-behaved language mapping mentioned in this corollary is one which essentially scrambles the images of words of different lengths in different ways.

**Example 3** Consider all permutations $\pi$ of the symbols of the alphabet $A$, where $\alpha > 2$. These can be enumerated, so that the set of all such permutations is $P = \{\pi_i : 0 \leq i < \alpha!\}$. Let $\pi_i$ also represent the extension of the permutation $\pi_i$ to a word morphism. Define language mapping $\Phi_P$ such that, if word $w$ is in language $L \in \mathcal{P}_A$ and $|w| = \alpha! \cdot q + r$, where $0 \leq r < \alpha!$, then (and only then) $\pi_i (w) \in \Phi_P (L)$. It can be seen that, although $\Phi_P$ is a nonerasing, nonexpansive, bijection,
but it is not a monoid morphism, for assume further that, for symbols \( a, b, c \in A \), permutation \( \pi_0 = \iota \), the identity, and permutations \( \pi_1 \) and \( \pi_{-1} = \pi_{\alpha!-1} \) are defined, in part by the following:

\[
\begin{align*}
\pi_1 : a & \mapsto b, \\
\pi_1 : b & \mapsto c,
\end{align*}
\]

\[
\begin{align*}
\pi_1 : c & \mapsto a, \\
\vdots
\end{align*}
\]

\[
\begin{align*}
\pi_{-1} : a & \mapsto b, \\
\pi_{-1} : b & \mapsto a, \\
\pi_{-1} : c & \mapsto c,
\end{align*}
\]

Of course, \( \Phi_P \) is injective and surjective. Let \( B = A \setminus \{a\} \). Consider the image of \( \Phi_P (B^{\alpha!-1}) \neq \Phi_P (B^{\alpha!-1}) \Phi_P (B^{\alpha!+1}) \). Note the symbol \( b \) cannot occur in any word in \( \Phi_P (B^{\alpha!-1}) \Phi_P (B^{\alpha!+1}) \), but \( B^{\alpha!-1} B^{\alpha!+1} = B^{2 \cdot \alpha!} \), and \( \Phi_P (B^{2 \cdot \alpha!}) = B^{2 \cdot \alpha!} \), and \( bbb \in B^{2 \cdot \alpha!} \). Therefore, \( \Phi_P \) is not a monoid morphism.

**Corollary 1.1.2** An injective language mapping \( \Phi \) which maps some language to a language of less cardinality is not a language morphism.

As another example of Lemma 1.1, namely, that it is not true that a language mapping is language code if and only if it is nonexpansive, we have the following.

**Example 4** Let the language mapping \( \Phi : \mathcal{P}_2 \to \mathcal{P}_4 \) be as follows:

\[
\begin{align*}
\Phi (\{\lambda\}) & = \{\lambda\} \\
\Phi (\{1\}) & = \{1, 3\} \\
\Phi (\{2\}) & = \{2, 4\}.
\end{align*}
\]

Then \( \Phi \) is clearly not nonexpansive, yet, since it is defined on each symbol of \( \mathcal{P}_2 \), it can be defined, by concatenation and set union, on all languages over \( \mathcal{P}_2 \). In this case, it is clear that the resulting
mapping is a language code, since, for language \( L \in \mathcal{P}_2 \) and word \( w \in \{1, 2\}^* \), \( \Phi (L) \) contains \( w \) if and only if \( w \) is in \( L \).

Now we can prove the rest: that only and necessarily a nonerasing, nonexpansive language morphism can be the extension of a word morphism, when this word morphism is itself code.

**Theorem 1.1** A language mapping \( \Phi : \mathcal{P}_A \to \mathcal{P}_B \) that is nonerasing and nonexpansive is a language morphism if and only if it is the extension to a language morphism of a code.

**Proof.** (\( \Rightarrow \)) Let \( \Phi \) be a nonexpansive, nonerasing language morphism from \( \mathcal{P}_A \) to \( \mathcal{P}_B \). Then, by Lemma 1.1, \( \Phi \) is a language code. By Fact 3, if word \( w \in A^* \), then \( \Phi (\{w\}) = \{\lambda\} \) if and only if \( w = \lambda \). For every symbol \( a \in A \), \( \Phi (\{a\}) \) is a singleton language \( \{w_a\} \) over \( B \), such that \( w_a = w_{a'} \) if and only if \( a = a' \). Suppose \( v \in A^k \), and \( k > 0 \). Then

\[
\Phi (\{v\}) = \Phi (\{v_1\} \{v_2\} \cdots \{v_k\}) = \Phi (\{v_1\}) \Phi (\{v_2\}) \cdots \Phi (\{v_k\}) = \{w_{v_1} w_{v_2} \cdots w_{v_k}\}. \tag{1.4}
\]

Define the literal mapping \( \phi : A \to B^* \) to be such that \( \phi (a) = w_a \) and let \( \phi \) also represent the extension of \( \phi \) to a word morphism. Let \( \Phi ' \) be the unique extension of \( \phi \) to a language morphism, as mentioned in Fact 4. Then \( \phi \) preserves concatenation of symbols, so that

\[
\Phi ' (\{v\}) = \Phi (\{v\}), \tag{1.5}
\]

by equation (1.4). But in addition, \( \phi \) is a code; for, suppose that \( \phi (u) = \phi (v) \). This means that, if \( |u| = k \) and \( |v| = l \), then

\[
\phi (u_1 \cdots u_k) = w_{u_1} \cdots w_{u_k} = w_{v_1} \cdots w_{v_l} = \phi (v_1 \cdots v_l)
\]

and thus that \( \Phi (\{u\}) = \Phi (\{u_1\} \cdots \{u_k\}) = \Phi (\{v_1\} \cdots \{v_l\}) = \Phi (\{v\}) \). But \( \Phi \) is a language code, so this is a contradiction, and so \( \phi \) is a code. Finally, by equation (1.3), \( \Phi ' \) preserves unions of words. Therefore, (1.2) holds for \( \Phi ' \). Since the same equation holds for \( \Phi \), we have that \( \Phi ' = \Phi \). 

18
We have seen that any word morphism \( \varphi \) extends to a language morphism, \( \Phi : \mathcal{P}_A \to \mathcal{P}_B \). If \( \varphi \) is also a code, then \( \# \Phi(F) = \# F \) for every finite language \( F \), and so \( \Phi \) is both nonexpansive and nonerasing.

Inspection of Example 2 reveals that the mapping \( \tilde{\Phi} \) is a language morphism that is not nonexpansive, which is why it maps \( \{ \lambda \} \) “poorly.” The following examples show a monoid morphism which is not nonerasing (and, therefore, is not the extension of any word morphism) and a word mapping which is not a code (and, therefore, extends to a language mapping which is not a nonexpansive, nonerasing language morphism).

**Example 5** Let \( \Phi \) be a language mapping such that, if \( L \in \mathcal{P}_A \), then

\[
\Phi(L) = \{ w \in L : x \in L \text{ implies } |x| \geq |w| \},
\]

i.e., \( \Phi(L) \) is the set of the shortest words in \( L \). Then \( \Phi(L) \subseteq A^{m_L} \), where \( m_L \) represents the length of the shortest word in \( L \neq \Lambda \), and \( \Phi(\Lambda) = \Lambda \). This is a monoid morphism, since, if \( L \) and \( M \) are non-empty languages, the shortest words in \( LM \), for \( L, M \in \mathcal{P}_A \) must be of the form \( uv \) where \( u \) is one of the shortest words in \( L \) and \( v \) is one of the shortest words in \( M \). Supposing that \( L = \Lambda \), then \( \Phi(L) = \Lambda \), in which case,

\[
\Phi(LM) = \Phi(ML) = \Lambda = \Phi(L) \Phi(M).
\]

The mapping \( \Phi \) is not non-erasing, since \( m_L \in \mathbb{N} \), and \( \Phi \), therefore, maps infinite languages into \( \text{FIN}_A \). Suppose that there exists a word morphism \( \varphi : A^* \to B^* \) such that \( \Phi(L) = \{ \varphi(w) : w \in L \} \) for every \( L \in \mathcal{P}_A \). Consider language \( L = \{ a, aa \} \in \mathcal{P}_A \). Then \( \varphi(a) = \varphi(aa) = a \). But \( \varphi \) is a word morphism, so \( \varphi(aa) = \varphi(a) \varphi(a) = aa \), a contradiction. Therefore, no such word morphism exists.

**Example 6** Let \( \sigma \), called the shift operator on languages, be defined thus:

\[
\sigma(w) = \begin{cases} 
  w_{[2,|w|]} & \text{if } |w| > 1, \text{ and} \\
  \lambda & \text{otherwise}.
\end{cases}
\]
Then $\sigma(L) = \{\sigma(w) : w \in L\}$. This operator produces a language in which all words have been “shifted to the left,” in that the symbol $w[i]$ in word $w$ of $L$ becomes symbol $i - 1$ in word $\sigma(w)$ of $\sigma(L)$, if $|w| > 1$, and, if $|w| = 1$ or $w = \lambda$, then $w$ gets “shifted” into $\lambda$. The shift operator preserves the identities $\{\lambda\}$ and $\Lambda$. But, for $w, v \in A^*$ such that $|w| > 1$ and $|v| > 1$, $|\sigma(w)\sigma(v)| = |\sigma(wv)| - 1$, so that $\sigma$ is not a word morphism. Nor is it a monoid morphism on a language space, as plainly $\sigma(\{w\} \{v\}) \neq \sigma(\{w\})\sigma(\{v\})$ if $w, v \in A^*$. Problems with the shift operator, as a language mapping, evidently originate from the fact that $\sigma$ is not nonerasing.

### 1.3.2 Language Space Isomorphisms

The final step in relating language spaces by a language mapping is to require surjectivity as well as injectivity. This bijection between two language spaces will be called an isomorphism. A semiring isomorphism is the most natural and most informative isomorphism.

**Definition 1.3.2** Two language spaces $P_A$ and $P_B$ will be said to be isomorphic if there exists a one-to-one monoid morphism between the two spaces, i.e., a bijective morphism

$$\Phi : P_A \rightarrow P_B$$

between $P_A$ and $P_B$. Then $\Phi$ will be called a monoid isomorphism. This will be denoted $P_A \cong P_B$. If $A = B$, then $\Phi$ is a monoid automorphism.

If $\Phi$ is a monoid isomorphism, i.e., $\Phi : P_A \rightarrow P_B$, then $\Phi$ will be called a natural language isomorphism if it is the extension to a language morphism of a word morphism $\varphi : A^* \rightarrow B^*$. Then the language spaces $P_A$ and $P_B$ will be said to be naturally isomorphic. If $A = B$, then $\Phi$ is a natural language automorphism.

The justification for this terminology lies in the fact that naturally isomorphic language spaces are, in essence, copies of each other. This is shown in the following fact and lemma.

**Fact 5** If $\Phi_{\varphi}$ is a natural language isomorphism, where $\varphi$ is the word morphism of which it is the extension, then $\varphi(\lambda) = \lambda$.

**Proof.** If $|\varphi(\lambda)| > 0$, then $|\varphi(a)| = |\varphi(a\lambda)| = |\varphi(a)\varphi(\lambda)| = |\varphi(a)| + |\varphi(\lambda)| > |\varphi(a)|$, which is impossible. □
Lemma 1.2 The following are equivalent:

1. the language spaces $\mathcal{P}_A$ and $\mathcal{P}_B$ are naturally isomorphic;

2. the alphabets $A$ and $B$ have the same cardinality;

3. there exists a surjective language code $\Phi$ from $\mathcal{P}_A$ onto $\mathcal{P}_B$; and

4. for a given permutation $\pi$ on the symbols of $A$, $\mathcal{P}_{\pi(A)}$ and $\mathcal{P}_B$ are naturally isomorphic.

Proof. (2$\Rightarrow$1) If $\#A = \#B$, then there exists a bijection $\varphi : A \rightarrow B$. Then by Lemma 1.1, the morphism $\varphi$, which is the extension to a code of literal mapping $\varphi$, extends to a nonerasing, nonexpansive language code $\Phi_\varphi : \mathcal{P}_A \rightarrow \mathcal{P}_B$. Note that $|\varphi(v)| = |v|$, for all $v \in \mathcal{P}_A$. Suppose word $w \in B^k$ for some $k \in \mathbb{N}$; if $k = 0$, $w = \lambda$, and $\Phi_\varphi(\{\lambda\}) = \{\lambda\}$ by Fact 3; if $k > 0$, then $w = w_1 \cdots w_k$, which is the $\varphi$-image of the unique word $\varphi^{-1}(w_1) \varphi^{-1}(w_2) \cdots \varphi^{-1}(w_k) \in A^k$, which is a word $v \in A^k$ such that $\Phi_\varphi(\{v\}) = \{w\}$, namely, the concatenation of the elements of $A$ which map, under $\varphi$, to $w_1, w_2, \ldots, w_k$, that is, the unique word $\varphi^{-1}(w) \in A^*$. From this observation, and by equation (1.2), $\Phi_\varphi$ is surjective. Consequently $\Phi_\varphi$ is a natural isomorphism between the language spaces $\mathcal{P}_A$ and $\mathcal{P}_B$.

(1$\Rightarrow$2) Assume there is a natural language isomorphism $\Phi_\varphi$ between $\mathcal{P}_A$ and $\mathcal{P}_B$, the extension of the morphism $\varphi : A^* \rightarrow B^*$. Then $\varphi$ which must be injective since $\Phi_\varphi$ is injective. For each $b \in B$ there exists a unique language $L_b \subseteq A^*$ such that $\Phi_\varphi(L_b) = \{b\}$. Suppose, for some $b \in B$, $w_b = uv$, where $u, v \in A^+$. Then $|\Phi(w_b)| = |\Phi(uv)| = |\Phi(u)\Phi(v)| = |\Phi(u)| + |\Phi(v)| = |b| = 1$, implying that $\Phi(u) = \lambda$ or $\Phi(v) = \lambda$, which is a contradiction. Therefore, $w_b \in A$ for each $b \in B$. Finally, $\Phi(A) \subseteq B$: indeed, if $a \in A$ and $\varphi(a) = t \in A^k$, with $k > 1$, then $\Phi_\varphi(\{a\}) = \Phi_\varphi(\{w_{t_1} \cdots w_{t_k}\})$, contradicting the bijectivity of a natural language space isomorphism. Therefore, the function $\tilde{\varphi} = \varphi|_A$ is a bijection between $A$ and $B$, and $\#A = \#B$.

(2$\Rightarrow$3) Trivial.

(3$\Rightarrow$2) If $\Phi$ is a surjective language code, then Lemma 1.1 says that $\Phi$ is a bijection between $\mathcal{P}_A$ and $\mathcal{P}_B$, and, therefore, by 1.1, the extension to a language code of a code $\varphi : A^* \rightarrow B^*$.

Further, the code $\varphi$ is surjective on $B^*$. For suppose, toward contradiction, that there is a word $w$ in $B^*$ without a pre-image in $A^*$. But this implies that the language $\{w\} \in \mathcal{P}_B$ has no pre-image
in $\mathcal{P}_A$, contrary to the hypothesis that $\Phi$ is surjective. For suppose that $\Phi (L) = \{ w \}$. If $v_1 \neq v_2$ and $v_1, v_2 \in L$, then $\Phi (\{v_1\} \cup \{v_2\} \cup \cdots ) = \{ w \}$, so $\Phi (\{v_1\}) = \Phi (\{v_2\}) = \cdots = \{ w \}$, since $\Phi$ is a language morphism, and this contradicts the injectivity of $\Phi$. It follows that $L$ is a singleton, but then, contrary to hypothesis, the single word in $L$ is the pre-image of $w$ under $\varphi$. Thus, $\Phi$ is indeed surjective.

If $b \in B$, then $\varphi^{-1} (b) \in A$ since $\varphi$ is a surjective code and if $|\varphi^{-1} (b)| > 1$, and $|\varphi^{-1} (b)| = uv$, where $u, v \in A^+$, then $\varphi [\varphi^{-1} (u) \varphi^{-1} (v)] = \varphi (w) = b$, contrary to the fact that $\varphi$ is a code. We need only show that $\varphi (A) \subseteq B$, and the proof is complete. Suppose toward contradiction that $|\varphi (a)| > 1$ for some $a \in A$, i.e., that $\varphi (a) = wv$ where $w, v \in B^+$. We then have that $\varphi (a) = \varphi (\varphi^{-1} (w) \varphi^{-1} (v))$ and $|\varphi^{-1} (w)| + |\varphi^{-1} (v)| > 1$ since $\varphi (\lambda) = \lambda$. But this says $\varphi$ is not a code, which is a contradiction. Hence, $\varphi (A) = B$.

(4$\Leftrightarrow$1) Every permutation $\pi$ is a bijection from $A$ onto $A$. Thus $\# \pi (A) = \alpha$. It has already been shown that $A$ and $B$ have the same cardinality if and only if $\mathcal{P}_A$ and $\mathcal{P}_B$ are naturally isomorphic. □

A natural consequence of this is the fact that automorphisms of $A$ lead to natural isomorphisms of $\mathcal{P}_A$, and that every aspect of a language space over $\alpha$ symbols is to be found in the canonical language space over $\alpha$ symbols.

**Corollary 1.2.1** The language space $\mathcal{P}_A$ is naturally isomorphic to the canonical language space over $\alpha$ symbols, and also to the language space $\mathcal{P}_{\pi (A)}$, where $\pi$ is a permutation of the symbols of $A$.

This result means that, unless a weighting is applied to the symbols of $A$, observations about language syntax are permutation-invariant. This, in turn, makes a permutation-invariant version of symmetric set difference useful.

1.3.3 The permutative set difference of languages

Permutations of the symbols of an alphabet extend to a natural automorphism of a language space. A variation of the symmetric set-difference follows from this, and extends from a straight-forward application to finite languages to a more intricate application to infinite languages.
**Definition 1.3.3** Enumerate the alphabet \( A \) as \( \{a_1, \ldots, a_\alpha\} \). This is a bijection between elements of \( A \) and \( \mathbb{N}_\alpha \), so it extends to a natural language isomorphism between \( \mathcal{P}_A \) and the canonical language space over \( \alpha \) symbols, \( \mathcal{P}_\alpha \). Let \( S_\alpha \) be the permutation group

\[
S_\alpha = \{ \pi \text{ such that } \pi : \mathbb{N}_\alpha \rightarrow \mathbb{N}_\alpha \}
\]

and let \( \pi \) also represent the extension of the permutation \( \pi \in S_\alpha \) to a language automorphism of \( \mathcal{P}_A \). For languages \( L \) and \( M \) in \( \mathcal{P}_A \) and a given \( \pi \in S_\alpha \), we will call the set \( L \triangle M \) the \( \pi \)-difference between \( L \) and \( M \), where \( L \triangle M \) denotes the set

\[
L \triangle M = \pi(L) \Delta M.
\]

Then \( L \triangle M \) will denote the \( \pi \)-difference \( L^{|<k|} \triangle M^{|<k|} \), and \( L \triangle M \) will denote the \( \pi \)-difference \( L^{|k|} \triangle M^{|k|} \).

The permutative set difference between finite languages \( F \) and \( G \) will be the minimum cardinality of the \( \pi \)-differences between \( F \) and \( G \) over all \( \pi \in S_\alpha \), and we will denote this quantity by

\[
\#(F \triangle G).
\]

In other words, \( \#(F \triangle G) \) is the cardinality of the smallest symmetric set difference between \( G \) and a permutation of symbols of the symbols of \( A \) applied to language \( F \). By Lemmas 1.1 and 1.2, this is the smallest cardinality of the set-difference between (naturally) isomorphic images of the two languages.

**Remark 1** Since \( \cdot \triangle \cdot \) is a group action on \( \mathcal{P}_A \times \mathcal{P}_A \), \( F \triangle G = G \triangle F \) and hence

\[
\#(F \triangle G) = \#(G \triangle F).
\]

The purpose of the permutative symmetric set-difference is to inspect structural differences between languages, rather than those brought about by a recoding of the symbols of a language. Let \( L \triangle M \) denote the \( \pi \)-difference \( L^{|<k|} \triangle M^{|<k|} \). Then

\[
\#(L \triangle M) = \#(L^{|k|} \triangle M^{|k|})
\]

if either \( L^{|<k|} \) or \( M^{|<k|} \) remains unaltered under permutations on the symbols of \( A \). We give a special term to such languages.

**Definition 1.3.4** If language \( L \in \mathcal{P}_A \), and \( \pi(L) = L \) for all \( \pi \in S_\alpha \), then \( L \) will be called \( \alpha \)-permutative. If the alphabet is understood, \( L \) will be called permutative. A pair of languages,
L, M ∈ ℰ_A, will be called relatively (α-)permutative if \( \# (L \triangle^k M) = \# (L \triangle^k M) \) for all \( \pi, \pi' \in S_\alpha \) and for all \( k \in \mathbb{N} \).

Examples of permutative languages are Λ, \( A^* \), and \( \{a^k_1, a^k_2, \ldots, a^k_\alpha\} \), where \( k \in \mathbb{N} \) and \( A = \{a_1, a_2, \ldots, a_\alpha\} \). In extending permutative set-difference to infinite languages, we need simply to take a limit on the \( \pi \)-differences between sections of the languages.

**Definition 1.3.5** The (general) permutative set-difference between languages L and M in ℰ_A is the limiting cardinality of the permutative set-difference between words of increasing length in L and M, denoted \( \# (L \triangle \alpha M) \). That is,

\[
\# (L \triangle \alpha M) = \limsup_{k \to \infty} \# (L \triangle^k \alpha M).
\]

The general permutative set-difference between two languages always exists, but is not necessarily given by a unique permutation in \( S_\alpha \). Definitions 1.3.3 and 1.3.5 obviously agree on finite languages.

1.4 Language pseudo-metrics and language norms

We now consider which characteristics are desirable in a function which maps each pair of languages to an unambiguous expression for the distance between them. Ideally, a language distance should be metric-like, for then the considerable accumulation of research regarding metric space would be applicable to the language space. But, in any case, we at least aim for a pseudo-metric language distance, where perhaps not every pair of distinct languages maps to a distance greater than zero. We also aim to incorporate characteristics which can be generally agreed to affect the distance between languages in an intuitive sense into the pseudo-metrics we use, provided these characteristics can be rigorously defined.

In subsection one, we assemble these definitions. The sole non-metric condition we actually impose is that two distinct languages should not be any closer to each other than a language formed from a subset of the words of one of them is from the other original language. We also define a naturally connected notion, which we call a language norm, which gives a meaning to the size of a language corresponding to the language pseudo-metric. In subsection two, we use a construction
borrowed from measure theory is used to show the bijective correspondence of language norms to language pseudo-metrics, and of permutative language norms to permutative versions of language pseudo-metrics.

1.4.1 Definition of language pseudo-metric and language norm

We are in pursuit of an adequate notion of a language pseudo-metric. Pseudo-metrics in general are defined as follows. Let $X$ be a space, and let $x, y, z,$ etc. represent elements of $X$. A pseudo-metric on $X$ is a function $\rho$ mapping pairs of elements of $X$ into the set of non-negative real numbers, under the conditions

\[
\rho(x, x) = 0, \text{ for all } x \in X, \tag{1.6}
\]

\[
\rho(x, y) = \rho(y, x), \text{ for all } x, y \in X, \text{ and } \tag{1.7}
\]

\[
\rho(x, z) \leq \rho(x, y) + \rho(y, z), \text{ for all } x, y, z \in X. \tag{1.8}
\]

Then the pair $(X, \rho)$ is called a pseudo-metric space.

The function $\rho$ is called a metric and $(X, \rho)$ is called a metric space if, for all $x, y \in X$, $x \neq y$ implies $\rho(x, y) > 0$. If, on the contrary, there exist $x, y \in X$ such that $x \neq y$ and $\rho(x, y) = 0$, $\rho$ we will call $\rho$ a strict pseudo-metric. If the image of $\rho$ is $\mathbb{R}^{\geq 0} \cup \{\infty\}$, then $\rho$ is a generalized pseudo-metric.

In our case, the space $X$ is a language space $\mathcal{P}_A$, the elements of which are languages. The most elementary notion of an increasing distance between languages is an increase in distinctions between distinct languages. We limit our consideration to, and specifically define as language pseudo-metrics, those functions which meet one additional condition.

**Definition 1.4.1** A pseudo-metric $d : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}^{\geq 0}$ will be called a language pseudo-metric on $\mathcal{P}_A$ if, for all languages $L, M, N \in \mathcal{P}_A$ such that $L \cap M = \emptyset$ and $N \subseteq M$,

\[
d(L, M) \geq d(L, N). \tag{1.9}
\]

The above condition says, in essence, that a language pseudo-metric on $\mathcal{P}_A$ preserves the partial order, under set inclusion, of the semiring $\mathcal{P}_A$. The cardinality of the symmetric set difference of
languages, for instance, is a generalized language pseudo-metric. From non-generalized pseudo-
metrics we can derive a different notion of the size of a language, a language norm.

**Definition 1.4.2** A language norm $\| \cdot \|$ on $\mathcal{P}_A$ is a function from the space of languages $\mathcal{P}_A$ into the non-negative real-numbers such that

$$\|\Lambda\| = 0, \text{ and}$$

L $\subseteq$ M $\subseteq$ $A^*$ implies $\|L\| \leq \|M\|$. (1.11)

Then $\|L\|$ will be called the norm of language L.

If the image of $\|\cdot\|$ is $\mathbb{R}^{\geq 0} \cup \{\infty\}$, then $\|\cdot\|$ will be called a generalized language norm.

If, in addition, for every language $L \in \mathcal{P}_A$, the norm is invariant on permutations of the symbols of the alphabet, i.e., $\|L\| = \|\pi(L)\|$ for any permutation $\pi \in S_\alpha$, then $\|\cdot\|$ will be called a permutative language norm.

Distinct language norms will be denoted with subscripts: $\|\cdot\|_\nu$, $\|\cdot\|_\xi$, etc., or $\|\cdot\|_a$, $\|\cdot\|_b$, etc.

Note that, if $\|A^*\| = 1$, and $\|\cdot\|$ is a language norm, then $\|\cdot\|$ is also a probability measure on $A^*$.

**Example 7** Cardinality of a language is a permutative generalized language norm. Cardinality is a measure on $A^*$, and so it fulfills the requirements of a language norm. Since the cardinality of words of any length is unchanged by a permutation of the symbols of $A$, it is permutative. Since there exist infinite languages, cardinality is a generalized permutative language norm.

Note that any subadditive measure and, therefore, any measure on $A^*$ is a language norm.

1.4.2 Link between language norms and language pseudo-metrics

From each language pseudo-metric there is a natural link to a unique language norm, and for each language norm there is a natural link to a unique language pseudo-metric. If a language norm is permutative, there is a link to a possibly distinct permutative language pseudo-metric. The construction involved in the proof of this depends, in part, on the following lemma from set theory.

**Lemma 1.3** For all sets $R$, $S$, and $T$, $R \Delta T \subseteq (R \Delta S) \cup (S \Delta T)$. 

26
For our purposes, we need to know that we can say something similar about permutative set-differences.

**Corollary 1.3.1** For finite languages $F$, $G$, and $H$ in $\mathcal{P}_A$, 

$$\# (F \triangle H) \leq \# (F \triangle G) + \# (G \triangle H). \quad (1.12)$$

**Proof.** Let permutation $\pi_1 \in S_\alpha$ be such that $\# \left( F \pi_1 \triangle H \right) = \# \left( F \triangle H \right)$. Let permutation $\pi_2 \in S_\alpha$ be such that $\# \left( H \pi_2 \triangle G \right) = \# \left( H \triangle G \right)$. By Remark 1, $\# \left( H \triangle G \right) = \# \left( G \triangle H \right)$. From Lemma 1.3,

$$\pi_1 (F) \triangle \pi_2 (H) \subseteq (\pi_1 (F) \triangle G) \cup (G \triangle \pi_2 (H)). \quad (1.13)$$

But, by Definition 1.3.3, $\# \left( F \pi_1 \right) = \# \left( F \triangle H \right) \leq \# \left( \pi_1 (F) \triangle \pi_2 (H) \right)$. At the same time,

$$\# \left[ \left( F \pi_1 \triangle G \right) \cup \left( G \pi_2 \triangle H \right) \right] = \# \left( F \pi_1 \triangle G \right) + \# \left( H \pi_2 \triangle G \right)$$

$$- \# \left[ \left( F \pi_1 \triangle G \right) \cap \left( H \pi_2 \triangle G \right) \right]$$

$$\leq \# \left( F \pi_1 \triangle G \right) + \# \left( H \pi_2 \triangle G \right)$$

$$= \# \left( F \triangle G \right) + \# \left( G \triangle H \right).$$

The conclusion follows. \[\square\]

This corollary establishes the general usefulness of permutative set-difference in the context of language pseudo-metrics. In particular, it shows that the triangle inequality holds for distances defined on permutative set-differences.

**Proposition 1** For each language norm $\| \cdot \| : \mathcal{P}_A \to \mathbb{R}_{\geq 0}$ there is a unique language pseudo-metric $d : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}_{\geq 0}$ such that $\| L \triangle M \| = d (L, M)$ for all $L, M \in \mathcal{P}_A$, and, for each language pseudo-metric $d : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}_{\geq 0}$, there is a unique language norm such that $\| L \| = d (L, \Lambda)$.

**Proof.** Uniqueness is obvious. Given language norms $\| \cdot \|_a$ and $\| \cdot \|_b$ on $\mathcal{P}_A$ such that $\| L \triangle M \|_a = d (L, M) = \| L \triangle M \|_b$ for all languages $L, M \in \mathcal{P}_A$, then

$$\| L \|_a = \| L \triangle \Lambda \|_a = d (L, \Lambda) = \| L \triangle \Lambda \|_b = \| L \|_b$$
and ∥·∥ₐ = ∥·∥ₖ. For any two language pseudo-metrics dₐ and dₖ on PA, if dₐ (L, M) = ∥LΔM∥ = dₖ (L, M) for all languages L, M ∈ PA and some language norm ∥·∥ on PA, then dₐ = dₖ.

Existence, then, remains to be shown.

If ∥·∥ₐ is a language norm on PA, then let dₐ be a function dₐ : PA × PA → ℝ₊₀ such that dₐ (L, M) = ∥LΔM∥ₐ. Then the function dₐ is well-defined, reflexive, and symmetric since the symmetric set-difference of languages L and M in PA is such that LΔM = MΔL and since, by the definition of dₐ, dₐ (L, L) = ∥Λ∥ₐ = 0. Applying Corollary 1.3.1, dₐ satisfies the triangle inequality as well. If L and M are disjoint languages and N ⊆ M, then ∥LΔN∥ₐ ≤ ∥LΔM∥ₐ by Property (1.11) of language norms, so dₐ (L, N) ≤ dₐ (L, M), and thus dₐ is a language pseudo-metric.

Now suppose that dₖ is a language pseudo-metric. Define ∥·∥ₖ : PA → ℝ₊₀ such that ∥L∥ₖ = dₖ (L, Λ). Since dₖ (L, L) = 0, ∥Λ∥ₖ = 0. If L ⊆ M, then M ∩ Λ = ∅, so dₖ (L, Λ) ≤ dₖ (M, Λ), so ∥L∥ₖ ≤ ∥M∥ₖ. Therefore, ∥·∥ₖ is a language norm on PA.

If, by use of the subscript ν, we distinguish a particular language norm ∥·∥ₐ over PA, then we will denote by dₐ, the associated language pseudo-metric, i.e., the language pseudo-metric agreeing with ∥·∥ₐ on symmetric set-differences, which has been shown to exist by Proposition 1. A similar association exists between permutative language norms and permutation-invariant distances, which we now show are also language pseudo-metrics.

**COROLLARY 1.3.2** For each permutative language norm ∥·∥ₐ : PA → ℝ₊₀, there is a permutative version of the associated language pseudo-metric dₐ, denoted by a superscript in parentheses indicating the number of symbols in A. In particular, the function dₐ(α) : PA × PA → ℝ₊₀, where

\[ dₐ(α) (L, M) = \min_{π ∈ S_α} ∥L Δ M∥ₐ, \]

is a language pseudo-metric.

**Proof.** Reflexivity and transitivity of the permutative version of a language pseudo-metric are established similarly to the way they were shown in the proof of Proposition 1. The triangle inequality is different. For if every π-difference between two languages is infinite, is it still possible that, for some π’, π” ∈ S_α, the norm of the π’-difference of the languages is less than the norm of their π”-difference. However, let L, M, and N be languages in PA, and suppose that π, π₁, and π₂ are the
permutations in $S_\alpha$ such that the norms $\| L \triangle N \|_\nu$, $\| L \triangle M \|_\nu$, and $\| M \triangle N \|_\nu$ are minimal over $S_\alpha$. Then, since $\| \cdot \|_\nu$ is a language norm, we have, by Remark 1 and Lemma 1.3, that

$$\| L \triangle N \|_\nu = \| \pi_1^{-1} (L) \triangle \pi_2^{-1} (N) \|_\nu \leq \| L \triangle M \|_\nu + \| M \triangle N \|_\nu,$$

as in the proof of Corollary 1.3.1. Since $\pi_2^{-1} \circ \pi_1$ is an element of $S_\alpha$ and since $\| \cdot \|_\nu$ is permutative, it follows that

$$\| L \triangle N \|_\nu \leq \| L \triangle M \|_\nu + \| M \triangle N \|_\nu.$$

This verifies the triangle inequality.

Finally, suppose that languages $L$ and $M$ are disjoint, and that language $N$ is a subset of language $M$. Since $\| \cdot \|_\nu$ is a language norm, $\| L \triangle N \|_\nu \leq \| L \triangle M \|_\nu$ for all $\pi \in S_\alpha$ (since $L \cap M = \Lambda$). But then, if $\| L \triangle N \|_\nu$ is minimal over $S_\alpha$, it cannot be more than $\| L \triangle M \|_\nu$ for any $\pi'$ in $S_\alpha$. Thus, $d_\nu^{(\alpha)}$ is a language pseudo-metric.

The notion of a series of elements converging to an element is a valuable tool for analysis. Convergence, however, requires that the limit of convergence be well-defined, and this may not be the case under a pseudo-metric, where some elements are at distance 0 from each other. In a metric space, convergence is commonly defined as follows.

**Definition 1.4.3** [Definition 3.1, Theorem 3.2[37]] Let $(X, \rho)$ be a metric space. The sequence $\{x_i\}_{i \in \mathbb{N}} \subseteq X$ is said to converge to $x$, denoted $x_i \to x$, if, provided, for all $\varepsilon > 0$, $\rho(x_i, x) < \varepsilon$ for all but finite $i \in \mathbb{N}$. Then $x$ is the limit of $\{x_i\}_{i \in \mathbb{N}}$.

1.5 Random languages

We now consider how to define a random language. A random element in a system is one which is typical of that system. In this way, a random element exemplifies the expressibility of a given structure. Since we have in mind, ultimately, application to systems of arbitrary variability, one justification for a given topology of languages is the characteristics of random languages under that topology. The idea is that these must exist and must be sufficiently indescribable to accommodate all heretofore undiscovered regularities in phenomena, but, at the same time should be hemmed in by languages that are describable, i.e., extraordinary enough that they may be described succinctly.
Historically, the description of such a characteristic of an element of a system is expressed by its complexity, in the sense of Kolmogorov complexity theory. There the Church-Turing thesis is employed to define complexity in terms of the bit-length of the shortest program which can generate an element of a system. Kolmogorov showed that this minimal length is independent of such things as the programming language employed, to within some constant of implementation. Nevertheless, the function \( K_a \), which gives the minimal length of the required to generate an element, is notoriously non-computable. In fact, random elements are precisely those elements \( x \in X \) such that

\[
K_a(x) \geq \ell(x),
\]

where \( \ell(x) \) is the bit-length of \( x \) itself.

It follows that the search for the “proof” that an element is random is a futile one. Such a proof in fact can generate the element using fewer bits than the element itself, and this is impossible.

Martin-Löf Randomness Tests first advanced in [29], are an outgrowth of Kolmogorov complexity theory. Instead of identifying random elements, a probability test is constructed which non-random elements fail with probability approaching one. In [29] it is shown that, under certain circumstances, there exists a universal randomness test, a way of specifying a region of the system which contains all the nonrandom elements.

The idea behind this is that nonrandom elements, being describable, must form a very small subset of the entire system. In the space of infinite binary sequences, for example, an open set may be understood as all sequences sharing an initial subsequence. Sequences which begin with many zeros are nonrandom, just as flipping a coin and coming up heads many times in a row is improbable. If one considers the nested sequence of open sets of infinite sequences such that, in the first open set, all sequences begin with zero, in the second, all sequences begin with two zeros, and so on, the intersection of the first \( k \) open sets in this sequence is all the sequences beginning with \( k \) zeros. It is reasonable to say that the measure of this intersection is \( 2^{-k} \). If a sequence is located withing this intersection, we can reject the hypothesis “this sequence is random” at a critical value of \( 2^{-k} \). Two things make this a test of randomness: first, the measure of the intersections of elements in this sequence diminishes, with each new set, by a factor of 2 and thus goes rapidly to zero; and, second, the nested intersection of these open sets is easily describable.
It is shown, in [29], that, in a topological space with a countable basis and a probability measure, randomness tests of this sort can always be constructed. What is more, it is shown in [29] that such a test satisfies all possible stochastic requirements of randomness, including requirements that have not yet been formulated. Finally, it is shown that, for infinite binary sequences, at any rate, a randomness test is devised such that, up to a fixed level of certainty, this test rejects all nonrandom elements in the system. Since the prevalence of randomness is one gauge of the richness of the information a system can express, it will be of significance to determine where the randomness lies in a language space topologized in a certain way. The following development of this concept comes from the work of Hertling and Weihrauch [16] and Calude, Marcus and Staiger[4]. We recall in this section the theorems and outline the process we mean to employ in evaluating the meaning of randomness under different language space topologies. For further details of the history and implications of the work on randomness, we refer the reader to texts such as Li and Vitanyi [25].

1.5.1 Martin-Löf randomness tests

In [4], the notions of a randomness space, a sequence of open sets computable from an enumerable basis, a randomness test, and a universal randomness test are defined, and we give these definitions here. Another necessary ingredient is an enumeration of $\mathbb{N} \times \mathbb{N}$. The enumeration of $\mathbb{N} \times \mathbb{N}$ adopted here will also be used in Chapter 2.

**Fact 6** The standard bijection $\langle \cdot, \cdot \rangle$ between $\mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$ will be made use of in this dissertation. The function $\langle \cdot, \cdot \rangle : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ assigns non-negative integers systematically to elements of $\mathbb{N} \times \mathbb{N}$. It assigns 0 to $(0, 0)$. Suppose all integers in the set $\{0, \ldots, n\}$ have been assigned, and that the number $n$ has been assigned to the pair $(i, j)$. If $i > 0$, $\langle \cdot, \cdot \rangle$ assigns the number $n + 1$ to the pair $(i - 1, j + 1)$. If $i = 0$, $\langle \cdot, \cdot \rangle$ assigns the number $n + 1$ to the pair $(j + 1, 0)$. In this way, assignment is made to all elements $(i, j)$ such that $i + j \leq k$ before any assignment is made to any pair $(i, j)$ such that $i + j > k$. Thus, $\langle \cdot, \cdot \rangle$ is injective. On the other hand, if $j = 0$, then $\langle i, j \rangle = \langle 0, i - 1 \rangle + 1$, or else $(i + 1, j - 1) \in \mathbb{N} \times \mathbb{N}$, and so $\langle i, j \rangle = \langle i + 1, j - 1 \rangle + 1$. Thus, the function $\langle \cdot, \cdot \rangle$ is a bijection. Calculation shows that this enumeration may be defined as follows:

$$\langle i, j \rangle = \frac{1}{2}(i + j)(i + j + 1) + j.$$
A randomness space, in short, is a space on which randomness tests exist. For this to be possible, we must be able to assign a measure to collections of elements of low complexity.

**Definition 1.5.1** [Calude, Marcus, Staiger [5]] A randomness space is a triple \((X, B, \mu)\) where \((X, \tau)\) is a separable topological space, \(\mu\) is a probability measure on the \(\sigma\)-algebra generated by \(\tau\), and \(B\) is the enumeration of a basis of \(\tau\).

We next define a means of effectively sorting out elements of \(X\) which may be labelled “non-random” at a specific confidence level. The confidence level is the \(\mu\)-measure of the intersection of a sequence of describable open sets. “Describability” is effectively approximated by the recursive enumerability of the numbers determining the successive open sets of this sequence as unions of basis elements of the topology. It is, of course, possible to describe all open sets as unions of basis elements. What we require beyond this is that the union of the sets of numbers corresponding to all basis elements comprising the the recursive enumeration of unions of open sets is possible because we have required that \(B\) is an enumeration of a basis of the topology \(\tau\).

**Definition 1.5.2** [Hertzling and Weihrauch, [16]] Let \((X, B, \mu)\) be a randomness space. Consider a sequence \(\mathcal{U} = \{U_i\}_{i \in \mathbb{N}} \subseteq \tau\). We say that \(\mathcal{U}\) is \(B\)-computable if there exists a recursively enumerable set \(N \subseteq \mathbb{N}\) such that

\[
U_i = \bigcup_{j \in N} B_j.
\]  

Then

1. A randomness test on \(X\) is a \(B\)-computable sequence \(\mathcal{V} = \{V_i\}_{i \in \mathbb{N}}\) such that \(\mu(V_i) \leq 2^{-i}\) for all \(i \in \mathbb{N}\).

2. An element \(x \in X\) is called nonrandom if \(x \in \bigcap_{i \in \mathbb{N}} V_i\) for some randomness test \(\{V_i\}_{i \in \mathbb{N}}\) on \(X\); an element of \(X\) is called random if it is not nonrandom.

The fact which makes Martin-Löf Randomness Tests so useful is that, under certain circumstances, a randomness space has a universal randomness test.

**Definition 1.5.3** [Martin-Löf, [29]] Let \((X, B, \mu)\) be a randomness space. Then a randomness test \(\mathcal{U} = \{U_i\}_{i \in \mathbb{N}}\) on \(X\) is universal if, for any randomness test \(\mathcal{V} = \{V_i\}_{i \in \mathbb{N}}\) on \(X\), there exists a fixed \(c \in \mathbb{N}\) such that \(V_{i+c} \subseteq U_i\) for all \(i \in \mathbb{N}\), where the constant \(c\) depends on \(\mathcal{U}\) and \(\mathcal{V}\), but not \(i\).
It is easy to see that this implies that every nonrandom element \( x \in X \) is an element of the intersection \( \bigcap_{i \in \mathbb{N}} U_i \). In general, verification of the existence of a universal randomness test requires a close examination of the randomness space. Certain results simplify the required examination. An \textit{upper-semi-computable} measure is one for which the set
\[
\{ W \in \tau : \mu(W) > 2^{-i} \}
\]
is recursively enumerable. In other words, an upper-semi-computable measure implies the existence of a Turing Machine which, given an open set in \( X \), generates a decreasing sequence of rational numbers converging to the measure of that open set.

\textbf{Theorem 1.2} [Theorem 3.10, [16]] Let \((X, B, \mu)\) be a randomness space. If \( \mu \) is upper-semi-computable, then there is a universal randomness test on \( X \).

A \textit{nowhere dense} set is one which contains no closure of an open set. A \textit{meager} subset of \( X \) is the union of countably many nowhere dense sets in \( X \). From Theorem 1.2 and Definition 1.5.2, we have the following.

\textbf{Theorem 1.3} [Theorem 3.11, [16]] If there is a universal randomness test \( U \) on the randomness space \((X, B, \mu)\), then

1. The set of nonrandom elements of \( X \) is of \( \mu \)-measure 0, and nowhere dense.
2. If the set of nonrandom elements of \( X \) is dense in \( X \), then the set of random elements is meager in \( X \).

Thus we have the well-know result that the properties of the nonrandom elements not found within a universal randomness test satisfy all stochastic requirements for randomness. Moreover, they satisfy every possible stochastic test for randomness, including any which have not been conceived.

1.5.2 A general approach to randomness in topological language spaces

The following is an outline, drawn from the definitions and theorems of the previous section, of the method we adopt for the examination of randomness in any given language space topology.
1. Elaborate the language space, or an appropriate topological space derived from it, as a separable space, and find an enumeration for a basis of the topology of that space.

2. Determine a measure on the $\sigma$-algebra generated by the language space (or other appropriate space) topology and prove its upper-semi-computability.

3. Display the associated nested open sets of languages, the measures of which converge quickly enough to yield a randomness test.

4. Draw conclusions about the character of a universal randomness test and, thereby, give a rough approximation to what a random language is under the given topology.

Thanks to the extensive work done on randomness, we will not need to fully elaborate this procedure in every case. For example, should we encounter a space homeomorphic to the unit interval, steps one, two, and three have long since been completed. The unit interval is mentioned as an example in [16], and we will make use of the principle conclusion mentioned there: constructible numbers, it is shown, are nonrandom.

With these preparatory definitions and observations, we turn to the consideration of three language pseudo-metrics. The first of these is a metric.
Chapter 2

Formal language space as a Cantor space

The metric set out in this chapter is shown by Genova and Jonoska in [13] and by Vianu in [38] to induce a homeomorphism between the space of formal languages and the Cantor space. We will therefore call it the Cantor distance on languages. A similar metric is applied to infinite sequences in [5, 25, 26], to biinfinite sequences in [5, 7, 23, 26] and to languages over infinite sequences in [5]. Before this similarity is explained, it is necessary to mention that, in [38], the space of formal languages topologized by Cantor distance is called the Bodnarchük metric space, and it is shown that there the Bodnarchük metric space space is homeomorphic to the learning space, under its usual metric. The Bodnarchük metric space however presupposes that the language space is transformed into a linear space over the field $\mathbb{F}_2$ (i.e., $(\{0, 1\}, +, \cdot)$). In this space, the Cantor language norm is an ordinary norm, where, for every language $L \in \mathcal{P}_A$, $0 \cdot L = \Lambda$ and $1 \cdot L = L$, and the sum of two languages is understood as their symmetric set-difference. A significant number of the conclusions presented here independently in [13] and [38], where formally the conclusions of the latter must be translated back from the normed linear space to the less-structured general language space. Vianu’s.

The related word metric $\omega : A^* \times A^* \to \mathbb{R}_{\geq 0}$ expresses the distance between words in the following manner. Let words $w, v \in A^*$. If $w = v$, then $\omega(w, v) = 0$. Otherwise, consider the languages $\text{Pref}(w)$ and $\text{Pref}(v)$. If $m_{w,v} = \min \{ k : \text{Pref}(w) \triangleleft^k \text{Pref}(v) \neq \Lambda \}$, then $\omega(w, v) = 2^{-m_{w,v}}$. That is to say, in establishing the $\omega$-distance between distinct $w$ and $v$, the entire suffix of each word beyond that initial distinction at the $m_{w,v}^{th}$ symbol is discarded. The metric $\omega$ can naturally be applied to infinite sequences. In like manner, the Cantor distance expresses the distance between two distinct languages using the shortest word-length at which differences between the two languages appear. Both $\omega$ and the Cantor distance we discuss discard two elements of the space in which they operate, but in the semiring of languages, this may an extremely large, extremely similar pair of languages. While there is no way to supplement distinct words with symbols to make them equivalent, there may very well be a way to add finitely many words to one language and make it
equivalent to another. It is therefore clear *a priori* that a language metric of this sort cannot express differences between families in the Chomsky hierarchy of languages.

In the first section, we define the Cantor language norm and language metric and present the standard basis elements of the metric topology induced by the Cantor distance. We show their relationship to the Bodnarchuk space. We also discuss the practical enumeration of the basis elements. The second section lists the facts about the Cantor topology which are significant for language spaces. In section three, Theorem 2.1 shows that, in the Cantor topological language space, the random languages are precisely the non-RE languages.

In Chapter 6, by the analogy to the topology developed by Calude, et al. and discussed in [5], we present a topology based upon the Cantor distance, but decisively different. In particular, under the derived topology, a language space is not necessarily homeomorphic to the Cantor space.

### 2.1 The Cantor language norm and distance

#### 2.1.1 Definition

In inspecting two languages, beginning with the shortest word in each language and proceeding to longer words, it seems natural to say that the sooner we find a distinction between the two languages, the farther apart they must be. From this point of view, only the word-length of the first observed distinction between the two languages is of significance. The language norm associated with such a concept of language distance is a function, which we denote \( \| \cdot \|_1 \), which assigns to each language the number \( \frac{1}{2} \) raised to the power of the length of the shortest word in the language.

**Definition 2.1.1** Let \( \| \cdot \|_1 : \mathcal{P}_A \rightarrow \mathbb{R}^{\geq 0} \) be the function such that, for language \( L \in \mathcal{P}_A \),

\[
\| L \|_1 = \begin{cases} 
0, & \text{if } L = \Lambda, \\
2^{-\min\{k : \#L[k] > 0\}}, & \text{otherwise}.
\end{cases}
\]

Thus, \( -\log_2 \| L \|_1 \in \mathbb{N} \) for all non-empty \( L \in \mathcal{P}_A \). For every pair of languages \( L, M \in \mathcal{P}_A \) such that \( L \subseteq M \), \( \| M \|_1 \) is no less than \( \| L \|_1 \), since \( L \cup M = M \). Therefore, the function \( \| \cdot \|_1 \) is a language norm. By Proposition 1, there is a unique pseudo-metric \( d_1 \) agreeing with \( \| \cdot \|_1 \) on the symmetric set-differences of languages.
**Definition 2.1.2** For language \( L \in \mathcal{P}_A \) and integer \( n \in \mathbb{N}, n > 1 \), define \( d_1 : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}_{\geq 0} \) as follows.

\[
d_1(L, M) = \|L \triangle M\|_1 = \begin{cases} 
0, & \text{if } L = M \\
2^{-\min \{k : L \triangle^k M \neq \emptyset\}} & \text{otherwise.}
\end{cases}
\]

Since \( d_1(L, M) = 0 \) only if \( L \triangle M = \emptyset \) and, therefore, only if \( L = M \), the language pseudo-metric \( d_1 \) is a language metric, which we will call the Cantor distance on languages.

Thus, the pair \((\mathcal{P}_A, d_1)\) is a metric space. As mentioned above, this can be made a linear space over \( F_2 \), and is then called the Bodnarchuk language space, normed by use of \( \|\cdot\|_1 \), and metrized by \( d_1 \) ([38]). Let \( \tau_1 \) be the metric topology induced on \( \mathcal{P}_A \) by the metric \( d_1 \). Then \((\mathcal{P}_A, \tau_1)\) will be called the Cantor topological space, and \( \tau_1 \) will be called the Cantor (language) topology, which equivalent to the topology of the Bodnarchuk metric space.

Note that the norm \( \|\cdot\|_1 \) is permutative: it makes no difference which symbols in the shortest words in a language. The permutative version of the metric \( d_1 \), however, may alter even though the permutative set difference is achieved by more than one permutation.

**Definition 2.1.3** For languages \( L \) and \( M \) in \( \mathcal{P}_A \), define \( d_1^{(\alpha)} : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}_{\geq 0} \) as follows:

\[
d_1^{(\alpha)}(L, M) = \begin{cases} 
0, & \text{if } L \triangle M = \Lambda \\
2^{-\min_{\pi \in S_\alpha} \left[ \min \{k : L \triangle^k M \} \right]}, & \text{otherwise.}
\end{cases}
\]

As an example, consider the languages \( L_1 = \{a, ab, aab, aaab, \ldots\} = aa^*b \cup \{a\} \) and \( L_2 = \{b, ba, baa, baaa, \ldots\} = ba^* \) over the alphabet \( \{a, b\} \). It is easy to see the permutative set-difference between \( L_1 \) and \( L_2 \) is not finite. The distance \( d_1(L, M) = \frac{1}{2} \), while \( d_1^{(\alpha)}(L_1, L_2) = \frac{1}{8} \), since, under the permutation \( \pi : a \mapsto b \) and \( \pi : b \mapsto a \), we find that

\[
\pi(L_1) \triangle L_2 = \{bba, baa, bbba, baaa, \ldots\},
\]

and, hence, that \( \|\pi(L_1) \triangle L_2\|_1 = \frac{1}{8} \).
With the Cantor distance or its permutative version, if two languages are each empty up to length $k$, and the cardinality of words of length $k$ differs in the two languages, then the distance between them is $2^{-k}$. The distance $d_1^{(\alpha)}$, however, unlike $d_1$, is a strict pseudo-metric. Consider a language $L \in \mathcal{P}_A$ which is not permutative. Then there is a permutation $\pi \in S_\alpha$ such that $\pi(L) \neq L$, and by the remarks following Definition 2.1.2, $d_1(L, \pi(L)) > 0$. But the distance $d_1^{(\alpha)}(L, \pi(L))$ is zero. This pseudo-metric, therefore, induces a topology on a language space which is not equivalent to the Cantor topology.

There may, however, be a quotient space of $\left(\mathcal{P}_A, d_1^{(\alpha)}\right)$ which is homeomorphic to the Cantor space. Further research is required to determine the exact character of the pseudo-metric topology induced by the permutative version of Cantor distance.

2.1.2 Language cylinder sets and their enumeration

By Corollary 2 of [38] there is an enumerable basis of open sets for the Bodnarchuk metric space. The open neighborhoods of radius $\varepsilon > 0$ around some language $L \in \mathcal{P}_A$, denoted $\mathcal{B}_\varepsilon(L) = \{M \in \mathcal{P}_A : d_1(L, M) < \varepsilon\}$, form the standard basis for $\tau_1$. Since distances between distinct languages are powers of $\frac{1}{2}$, it follows that elements of the standard metric basis of the Cantor topology form the collection

$$\mathcal{C} = \{\mathcal{B}_{2^{-k}}(L) : k \in \mathbb{N}, L \in \mathcal{P}_A\}. \quad (2.2)$$

**Remark 2** The set $\mathcal{B}_{2^{-k}}(L)$ is the collection of languages which agree with $L$ on all words of lengths up to and including length $k$.

Referring to the word-metric $\omega$, a cylinder set of words over $A$ is defined as a collection $C_w = w A^*$, for some word $w$ in $A^*$. Then $C_w$ is the set of all words $v$ such that $w \in \text{Pref}(v)$. We see that, analogously, the basis elements of $\tau_1$ are the cylinder sets of languages, i.e., all the languages in $\mathcal{P}_A$ which agree with some given language $L$ on all words of all lengths up to and including some length $k$.

**Definition 2.1.4** [Vianu, Daniela[13, 38]] The language cylinder set of length $k \in \mathbb{N}$, denoted $C_{L,k}$, around language $L \in \mathcal{P}_A$ is the set

$$C_{L,k} = \{M \in \mathcal{P}_A : L[\leq k] = M[\leq k]\}. \quad (2.3)$$
REMARK 3 Given $k \in \mathbb{N}$, each distinct subset of $A^{\leq k}$ corresponds to a single cylinder set of languages of length $k$. Note that $C_{L, k+1} \subseteq C_{L, k}$, for the language $M = (L \cap A^{\leq k}) \cup (L^c \cap A^{k+1})$ is non-empty for all $L \in \mathcal{P}_A$ and all $k \in \mathbb{N}$. Yet $M \in C_{L, k} \setminus C_{L, k+1}$.

REMARK 4 The intersection of two cylinder sets, say $C_{L, j}$ and $C_{M, k}$, where $j \leq k$, is either $\emptyset$ or $C_{M, k}$. To be precise,

$$C_{L, j} \cap C_{M, k} = \begin{cases} \emptyset, & \text{if } L^{[\leq j]} \neq M^{[\leq j]} \\ C_{M, k}, & \text{if } L^{[\leq j]} = M^{[\leq j]} \end{cases}.$$ 

For, suppose language $N \in C_{L, j} \cap C_{M, k}$ and $L^{[\leq j]} \neq M^{[\leq j]}$. Then there exists $i \leq j$ such that $L[i] \neq M[i]$. But we have, by our supposition, that $N[i] = L[i]$ and $N[i] = M[i]$, which is impossible.

In addition, note that the intersection of countably many cylinder sets contains a single language, and is not a cylinder set. In particular,

$$\bigcap_{i \in \mathbb{N}} C_{L, i} = \{L\},$$

which is not a cylinder set.

REMARK 5 The union of two cylinder sets, $C_{L, j} \cup C_{M, k}$, where $j \leq k$ is either a single cylinder set or a union of disjoint cylinder sets. To be precise,

$$C_{L, j} \cup C_{M, k} = \begin{cases} C_{L, j}, & \text{if } L^{[\leq j]} = M^{[\leq j]} \\ C_{L \setminus M^{[\leq j]}, j} \cup C_{L \cap M^{[\leq j]}, j} \cup C_{M^{[\leq j]}, k}, & \text{if } L^{[\leq j]} \neq M^{[\leq j]} \end{cases}.$$ \hspace{1cm} (2.4)

where the last case is the union of three disjoint cylinder sets, two of length $j$, the third of length $k$. Since this fragmentation into disjoint cylinder sets can continue indefinitely, a countable union of cylinder sets is a countable union of disjoint cylinder sets.

Let $\mathcal{C}_k$ denote the collection of all language cylinder sets of length $k$. That is, let

$$\mathcal{C}_k = \{C_{L, k} : L \in \mathcal{P}_A\}.$$
From (2.2) and (2.3), we have that $C = \bigcup_{k \in \mathbb{N}} C_k$, i.e., that the collection of language cylinder sets is the standard basis for the Cantor topology.

The discussion of randomness on the space $(\mathcal{P}_A, \tau_1)$ requires an enumeration of $C$, i.e., a bijection $E : \mathbb{N} \mapsto C$. We construct $E$ using the fact that each cylinder set is defined by a finite language. For some enumeration of $A$, $a : \mathbb{N}_\alpha \mapsto A$, define the length-lexical order $\prec$ on $A^*$ as follows. If word $w, u \in A^*$ then
\[
  w \prec u \text{ if } \begin{cases} 
  |w| < |u| \\
  |w| = |u| \text{ and } w = xa_iy, u = xa_jz, \text{ where } x, y, z \in A^* \text{ and } i < j. 
\end{cases}
\]

Set $l(\lambda) = 0$ and let $l(w) = l(u) + 1$ if $v \prec u$ implies $v \prec w$ or $v = w$. The function $l : A^* \mapsto \mathbb{N}$, called the length-lexical enumeration of $A^*$, is a bijection.

Then $l$ induces an enumeration $\tilde{l}$ of $\text{FIN}_A$. Specifically, if $F \in \text{FIN}_A$, let
\[
  \tilde{l}(F) = \sum_{w \in F} 2^{l(w)}.
\]

Thus, $\tilde{l}$ maps $\text{FIN}_A$ into $\mathbb{N}$. Since $l$ is bijective, so, too is $\tilde{l}$. Every nonnegative integer has a unique binary expansion, meaning that every number $n$ is the image, under $\tilde{l}$, of one and only one finite language. Recall the enumeration $\langle \cdot, \cdot \rangle$, shown to be a bijection between $\mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$, and consider the mapping $E : \mathbb{N} \mapsto C$, where
\[
  E^{-1}(C_{F,k}) = \left\{ \tilde{l}(F), k - \max \{ |w| : w \in F \} \right\}.
\]

Since every distinct cylinder set $C_{L,k}$ is determined by the finite language $L \cap A^{\leq k}$, there is one distinct language cylinder sets of length $k$ for each distinct finite language $F$ such that $F \subseteq A^{\leq k}$. Moreover, for every finite language $F \subseteq A^{\leq k}$ there is a distinct language cylinder set of length $m$ for each integer $m \geq k$. Thus, corresponding to each non-negative value of the expression
\[
  k - \max \{ |w| : w \in F \},
\]
there is exactly one cylinder set in \( \mathcal{C}_k \), by Remark 3. As \( \tilde{l} \) has been shown to be a bijection, the arguments of \( \langle \cdot, \cdot \rangle \) in (2.5) take on every non-negative value. By the definition of cylinder sets, the mapping \( E \) is surjective. Thus, \( E \) is an enumeration of the basis elements of the Cantor topology. Set \( E_j = E(j) \).

2.2 The Cantor topology on a language space

The main topological features of \( (\mathcal{P}_A, \tau_1) \) are shown in [38] and [13]. They are presented here without proof.

**Lemma 2.1** In \( (\mathcal{P}_A, \tau_1) \), every cylinder set is both closed and open.

Convergence in \( (\mathcal{P}_A, d_1) \) may be described as follows.

**Lemma 2.2** A sequence of languages \( \{L_i\}_{i \in \mathbb{N}} \subseteq \mathcal{P}_A \) converges to the language \( L \in \mathcal{P}_A \) in \( (\mathcal{P}_A, \tau_1) \) if and only if, for all \( m \in \mathbb{N} \), \# \( (L_i \Delta^m L) \) = 0 for all but finitely many \( i \in \mathbb{N} \). In this case, we write \( L_i \rightarrow L \).

This simply says that a language sequence converges if, for any \( m \in \mathbb{N} \), all but finitely many languages in the sequence agree on all words of length less than \( m \). From this and the observation that \( L_i^{[i]} \rightarrow L \) we have the following.

**Corollary 2.2.1** The finite languages are dense in a space of languages under the \( \tau_1 \) topology.

**Lemma 2.3** The topological space \( (\mathcal{P}_A, \tau_1) \) is homeomorphic to the Cantor space.

**Proof.** The length-lexical enumeration \( l \) induces a bijection \( f \) between the language space \( \mathcal{P}_A \) and \( \{0, 1\}^\omega \), the space of infinite sequences over the alphabet \( \{0, 1\} \), where \( f : L \mapsto x \) such that \( x_{l(w)} = 1 \) if and only if \( w \in L \). Then \( f \) is a homeomorphism, as shown in [13], and the space of binary sequence is bijective with the ternary Cantor set. \( \square \)

**Corollary 2.3.1** \( (\mathcal{P}_A, \tau_1) \) is compact, perfect, and totally disconnected.

**Proof.** These are the well-known features of the Cantor space [30]. \( \square \)

In addition, there is both an RE language and a non-RE language in every open set of \( (\mathcal{P}_A, \tau_1) \). This, together with Corollary 2.2.1, means that there is no distinction under the Cantor topology between any of the classical families of languages.
2.3 Randomness under Cantor distance

The question “what is a random language?” has a specific answer under the topology induced by Cantor distance. Namely, we show that the terms “random” and “non-recursively enumerable” are interchangeable when describing languages in the Cantor topological space \((\mathcal{P}_A, \tau_1)\). We following the outline proposed in Subsection 1.5.2.

We have established the first point in the outline already. An enumeration of a countable basis for the space has been established: the function \(E\) introduced in Subsection 2.1.2 enumerates the standard metric basis of \(\tau_1\). In the first subsection below, we introduce the second element: a probability measure on the \(\sigma\)-algebra generated by the topology \(\tau_1\). This measure is shown to be upper-semi-computable in Lemma 2.4. In subsection two, we exhibit a randomness test, the characteristics of a universal randomness test, and examples of nonrandom elements, showing why they are nonrandom. Finally, in the subsection three, we draw conclusions about random and nonrandom elements of a language space topologized by the Cantor distance. The equivalence of randomness and non-RE is demonstrated in Theorem 2.1.

2.3.1 An upper-semi-computable measure on open sets

The following construction of a measure on the \(\sigma\)-algebra induced by \(\tau_1\) on the language space is analogous to the Bernoulli process. The Bernoulli process assigns a measure to a word as the product of a probability measure on an alphabet, as discussed, for instance, in [26]. Our construction must differ from this slightly, because a language space is not the product of languages of different lengths, as we have emphasized already. The number of possible languages containing words of a maximum length \(l\) is of the order \(O\left(2^{\alpha l+1}\right)\). This means that, to obtain a uniform probability measure on language cylinder sets of length \(l\), we require an expression of the form \(\exp\left[-\exp(l)\right]\).

Under the Bernoulli process, a measure \(\mu_B(a) = 1/\alpha\) is assigned to each symbol \(a\) of the alphabet \(A\). The measure \(\mu_B(w) = \prod_{i=1}^{|w|} \mu_B(w[i]) = \alpha^{-|w|}\) is assigned to word \(w \in A^*\). Consider the elements of the \(\sigma\)-algebra generated by \(\tau_1\), viz., the unions of distinct elements of \(\mathcal{C}\). Observe that \(|\mathcal{C}_k| = |\mathcal{P}(A^k)| \cdot |\mathcal{C}_{k-1}|\). By induction (see also Proposition 3, [38]),

\[
|\mathcal{C}_k| = \prod_{i=0}^{k} 2^{\alpha^i} = 2^{\frac{\alpha^{k+1} - 1}{\alpha - 1}}.
\]
Definition 2.3.1 Let the measure $\mu$ on the $\sigma$-algebra generated by $\tau_1$ on $\mathcal{P}_A$ to be the function such that

$$\mu(C_{L,k}) = \frac{1}{\#C_k} = 2^{\frac{k+1}{k-1}}.$$ 

Then, first, $\mu$ is a probability measure, since the measure of all disjoint cylinder sets of length $k$ in $\mathcal{C}_k$ sums to 1, and this union is exactly $\mathcal{P}_A$. Secondly, we have the enumeration $E$ of basis elements of $\tau_1$, defined in equation (2.5). It follows, from Definition 1.5.1 on page 32, that $(\mathcal{P}_A, E, \mu)$ is a randomness space. A randomness test on $(\mathcal{P}_A, E, \mu)$, by 1.5.2, is a sequence of open sets $\{L_i\}_{i \in \mathbb{N}}$ in $\mathcal{P}_A$ such that $\mu(L_i) \leq 2^{-i}$ and such that, for some recursively enumerable set $N \subseteq \mathbb{N}$

$$L_i = \bigcup_{j \in \mathbb{N}} E_{i,j}$$

i.e., $\{L_i\}_{i \in \mathbb{N}}$ is an $E$-computable sequence.

Our next concern is to verify that a universal randomness test exists on $(\mathcal{P}_A, E, \mu)$. To clarify the following proof, let the representative of $C_{L,k}$ of length $m$, where $m \geq k$, be the collection of languages in $\mathcal{P}(A^{\leq m})$ which are in $C_{L,k}$. If $k = m$, this set is a singleton, namely, the language $L^{[\leq k]}$. We will call this the representative of $C_{L,k}$.

Lemma 2.4 The measure $\mu$ (from Definition 2.3.1) is upper-semi-computable.

Proof. Proceed by an enumerated process of inspection of an open set. Let $L$ be a countable union of language cylinder sets. At stage $i$, let the language set $L_i$ signify the set $L_i = \{L^{[\leq i]} : L \in L\}$, that is, all languages of made up of all the words of length no greater than $i$ in some language in $L$. Then $L_i$ is a finite set. Let $\mathcal{C}_{L,i}$ be the collection of the disjoint cylinder sets up to word length $i$, represented in each $L_i$, for $0 \leq i' \leq i$. Then, in symbols,

$$\mathcal{C}_{L,i} = \{C_{L,i} : 0 \leq j \leq \#L_i, L_j \in L_i\},$$

i.e., $\mathcal{C}_{L,i}$ is the collection of all distinct cylinder sets of length $i$ represented in the collection $L_i$. If, for convenience and clarity, we write $C_{L,i-1} \in \mathcal{C}_{L,i}$ when $L \subseteq A^{<k}$ and $L \cdot \mathcal{P}(A^k) \subseteq L_i$, and so write $C_{L,i-j} \in \mathcal{C}_{L,i}$, for $1 < j \leq i$, if $C_{LM,i} \in \mathcal{C}_{L,i}$, for each $M \in \mathcal{P}(A^{i-j+1})$, we arrive at an
equivalent definition of $C_{L,i}$, namely

$$C_{L,i} = \left\{ C_{M,j,m,k} : 0 \leq j \leq i', \{ k_m \}_{m=0}^{m''} \subseteq \mathbb{N}_i \text{, for some } i' \leq \#C_{km} \text{ and some } i'' \leq i \right\}, \tag{2.7}$$

that is, $C_{L,i}$ is the set of distinct language cylinder sets of length $i$ found in $L_i$, as in 2.6, which can be equivalently written as a collection of language cylinder sets of minimal length such that all of their representatives of length $i$ are found in $L_i$, as in 2.7.

To clarify further, suppose that the alphabet is $A = \{ a, b \}$ and that $L$ is some open set in $(\mathcal{P}_A, \tau_1)$. Suppose that all languages in $L$ are $\lambda$-free, which means that $C_{L,0} = \{ C_{\lambda,0} \}$. Suppose that, at length 1, we find that every language in $L$ contains either the word $a$ or the word $b$, but not both. Then $L_1 = \{ \{ a \}, \{ b \} \}$, and $C_{L,1} = \{ C_{\{ a \}, 1}, C_{\{ b \}, 1} \}$. Note that both cylinder sets in $C_{L,1}$ are subsets of the only element of $C_{L,0}$, but that there are other cylinder sets of length 1 which are subsets of $C_{\lambda,0}$.

Suppose that, at length 2, we find the language collection $L_2 = \{ \{ a, aa, ab, ba, bb \}, \{ a, aa, ab, ba \}, \{ a, aa, ba, bb \}, \{ a, ab, ba, bb \}, \{ a, aa, ab \}, \{ a, aa, ba \}, \{ a, aa, bb \}, \{ a, ab, bb \}, \{ a, ab \}, \{ a, ba \}, \{ a, bb \}, \{ a, bb \}, \{ b, aa \}, \{ b, bb \} \}$.

Then $C_{L,2} = \{ C_{\{ a \}, 1}, C_{\{ b, aa \}, 2}, C_{\{ b, bb \}, 2} \}$, because every representative of $C_{\{ a \}, 1}$ of length 2 is a language in $L_2$, but only two cylinder sets in $C_{\{ b \}, 1}$ of length 2 are included. Of course, we could write this in the manner of (2.6), but it we see that the form specified in (2.7) is much simpler.

Moreover, in the third step of the procedure, we need only look for language which are subsets of the cylinder sets in $C_{L,2}$.

After each step of this procedure, the $\mu$-measure of $L$ is bound by the sum of the measure of the cylinder sets in $C_{L,i}$:

$$\mu(L) \leq \sum_{j=0}^{\#L_i} \mu(C_{L,j,i}) = \frac{\#L_i}{\#C_i}$$

$$= \sum_{m=0}^{i''} \sum_{j=0}^{i'_m} \frac{1}{\#C_{km}} = \sum_{m=0}^{i''} \frac{i'_m}{\#C_{km}}.$$
The initial step in this procedure, step 0, is checking whether every language in \( L \) is \( \lambda \)-free. Then \( L_0 \subseteq \{ \{ \Lambda \}, \{ \lambda \} \} \) and \( L_0 \) is either \( \{ C_{A,0}, \{ \lambda \}, 0 \} \), or \( \{ C_{A,0}, C_{(\lambda),0} \} \). We identify, in the next step, step 1, the collection \( A \subseteq A^{\leq 1} \) of subsets of \( A \cup A^0 \) found as the subsets of languages in \( L \) of words of lengths less than 2, and label this set \( L_1 \).

Assuming the above procedure has been followed and at step \( k \) we have available the finite sets \( C_{L,k-1} \) and \( L_k \), we proceed as follows: for each cylinder set \( C_{M,j,m} \) in \( C_{L,k-1} \), count the distinct languages \( L \) in \( L_k \) such that \( M_{j,m} \subseteq L \). Say that there are \( c \) such languages \( L \) in \( L_i \). If \( c < \#C_k / \#C_{k-1} = 2^{\alpha_k} \), then place in \( C_{L,k} \) the cylinder sets \( C_{L,k} \), \( L \in L_k \) such that \( M_{j,m} \subseteq L \). If \( c = 2^{\alpha_k} \), place \( C_{M_{j,k}, j} \) in \( C_{L,k} \). This is a recursively enumerable procedure for each \( k \). If \( L \) is the union of countably many open sets in the topological space \( ( \mathcal{P}_A, \tau_1 ) \), this procedure is recursively enumerable. The sequence

\[
\left\{ \sum_{j=1}^{\#L_i} \mu \left( C_{L,j,i} \right) \right\}_{i \in \mathbb{N}}
\]

approximates the measure of \( L \) from above. Hence, \( \mu \) is upper semi-computable.

In the example given in the body of the proof, our approximation to the measure of \( L \) would result, at step 0, in \( \frac{1}{2} \), since \( \#C_0 = 2^{d+1-1} = 2 \). At step 1, the our approximation to the measure of \( L \) is \( 1/8 + 1/8 = 1/4 \), since we found two cylinder sets of length 1. At the second step, we have one cylinder set of length 2 and two of length 2, so our second approximation is \( 1/8 + 1/2 + 1/2 = 9/64 \).

**Corollary 2.4.1** There exists a universal randomness test \( \mathcal{U} = \{ U_i \}_{i \in \mathbb{N}} \) on \( ( \mathcal{P}_A, \mathcal{E}, \mu ) \).

**Proof.** By Theorem 1.2 on page 33: we have shown that \( ( \mathcal{P}_A, \mathcal{E}, \mu ) \) is a randomness space. Lemma 2.4 shows that \( \mu \) is upper-semi-computable, fulfilling the conditions of Theorem 1.2. Therefore, there exists a universal randomness test \( \mathcal{U} = \{ U_i \}_{i \in \mathbb{N}} \) on \( ( \mathcal{P}_A, \mathcal{E}, \mu ) \), meaning that nonrandom element \( L \), labelled nonrandom by some randomness test \( \mathcal{V} \) at a certainty level of \( 2^{-i} \), is labelled nonrandom by \( \mathcal{U} \) with an uncertainty of at most \( 2^{-i+j} \), where \( j \) is only dependent on \( \mathcal{U} \) and \( \mathcal{V} \), but not \( L \). of the space fails the test, and is found within the intersection

\[
\bigcap_{i \in \mathbb{N}} U_i.
\]

\( \square \)

45
The concept of representative, as used in the proof of Lemma 2.4, makes it clear how it is possible that a universal randomness test for languages might exist. The definition of a randomness test is that $\mu(U_i) < 2^{-i}$. This would seem at first to imply that, over an alphabet with two symbols, a universal randomness test would involve a choice of four out of the eight cylinder sets of length 1, contrary to the intuition that any finite language is nonrandom (which is, indeed, the case, as shown below). Instead, of course, the universal randomness test includes only much longer cylinder sets. For example, by including a maximum length of just 3, there are $32,768$ cylinder sets over an alphabet of two symbols, and, if only cylinder sets of this length occurred in $U_1$, there could be $16,384$ cylinder sets in $U_1$, easily accommodating a representative of length 3 of every cylinder set of length one!

2.3.2 Examples of nonrandom languages

We consider the most obvious instances of languages which should be nonrandom, and show that they belong to randomness tests.

**Example 8** $A^*$ is nonrandom. Recalling the enumeration $\tilde{I}$ of finite languages, the sequence

$T = \{t_i\}_{i \in \mathbb{N}} = \left\{ \left\langle i, \tilde{I}(A^{\leq i}), 0 \right\rangle \right\}_{i \in \mathbb{N}}$

is recursively enumerable. In fact, $\tilde{I}(A^{\leq i}) = 2^{2^i - 2^{-i-1}} - 2^{2^{-i-1}}$ for $i \in \mathbb{N}$. The sequence computed from $T$, namely $\mathcal{T} = \{E_i\}_{i \in \mathbb{N}}$, is precisely the sequence $C_{\{\lambda\}, 0}$, $C_{\Lambda \cup \{\lambda\}, 1}$, $C_{\Lambda \cup \{\lambda\}, 2}$, ... The intersection of these cylinder sets is equal to $A^*$, and any other language is at a positive distance from all languages in all but finitely many cylinder sets in $\mathcal{T}$.

**Example 9** Any finite language $F$ is nonrandom, as there is a maximum length $n \in \mathbb{N}$ of a word in $F$. Therefore, $F$ is the sole language found in the intersection of the cylinder sets

$C_{F \cap A^{\leq a}, 0}, C_{F \cap A^{\leq a} + 1}, \ldots, C_{F \cap A^{< n}, n}, C_{F, n}, C_{F, n + 1}, C_{F, n + 2}, \ldots \tag{2.8}$

and the sequence enumerating these cylinder sets is clearly recursively enumerable.

**Example 10** The language $a^*$, for any fixed $a \in A$, is nonrandom; for suppose that under a length-lexical ordering $l$ of $A^*$, $l(a) = 1$. Then consider the sequence $V = \{v_i\}_{i \in \mathbb{N}}$, where $v_1 =$
\( i, \left\langle \sum_{j=0}^{i} \sum_{k=0}^{j} 2^{a_k}, 0 \right\rangle \). For word length \( i \), this selects the language \( L_i = \{ \lambda, a, a^2, \ldots, a^i \} \) and returns the number associated by \( E \) with the language cylinder of length \( i \) set around \( L_i \). Then \( C_E v_i, i = C_{L_i, i} \), and \( a^* \) is the unique language in the following intersection of cylinder sets:

\[
\bigcap_{i \in \mathbb{N}, (i, j) \in V} E_j.
\]

It may seem that any language we can describe effectively is nonrandom; for, having an effective way of describing the words of length \( k \), for all \( k \in \mathbb{N} \), seems to imply that the bijection in the enumeration of the cylinder sets will allow us to set up the appropriate randomness test to accommodate this language, and no other. This is, in fact, not only the gist of the Martin-Löf randomness test notion, but explicitly the case under the Cantor topology.

2.3.3 Random equals non-recursively enumerable

We now identify the random languages as the co-nowhere dense set of languages which can be defined by syntax-grammars. The meaning of this languages we cannot hope to describe or generate by Turing Machines, and only such languages, can be considered random in the Cantor language topology.

**Theorem 2.1** Language \( L \in \mathcal{P}_A \) is nonrandom in \( (\mathcal{P}_A, \tau_1) \) if and only if it is an RE language. The set of random languages is of measure 1 and meagre in the randomness space \( (\mathcal{P}_A, E, \mu) \). The set \( RE_A \) is of measure 0 and nowhere dense in the randomness space \( (\mathcal{P}_A, E, \mu) \).

**Proof.** By Corollary 2.4.1, there exists a universal randomness test \( \mathcal{U} = \{ U_i \}_{i \in \mathbb{N}} \) on \( (\mathcal{P}_A, E, \mu) \), \( E \) is the enumeration of cylinder sets from above and \( \mu \) is defined by its value on cylinders, so that \( \mu (C_{F, i}) = 2^{-\frac{a^2-1}{n-1}} \). It follows from Theorem 1.3 on page 33 that the measure of the set of random languages is \( \mu (\mathcal{P}_A) - 0 = 1 \). As \( n \to \infty \), the closure of every open set is excluded from the intersection \( \bigcap_{i=0}^{n} U_i \), since \( \mu (U_i) < 2^{-i} \) for all \( i \in \mathbb{N} \). It remains to be shown that a language is nonrandom if and only if it is RE.

Consider a universal randomness test \( \mathcal{U} \) and the sequence \( N = \{ n_i \}_{i \in \mathbb{N}} = \left\{ \frac{a^i-1}{a-1} + 1 \right\}_{i \in \mathbb{N}} \). Then, for each \( i \in \mathbb{N} \), \( \mu (U_{n_i}) < 2^{-\frac{a^i-1}{a-1}+1} < \mu (C_{F, i}) \) for any \( C_{F, i} \in C_i \), meaning that no cylinder set of length \( i \) (or less) is contained in \( U_{n_i} \). Therefore, the words of length up to \( i \) of
any language in the intersection $\bigcap_{j=0}^{n_i} U_j$ are computed for every language in $\mathcal{U}$ by the recursively enumerable sequence $\{N_j\}_{j=0}^{n_i}$ which generates the open sets $U_0, \ldots, U_{n_i}$. The sequence

$$\{N_j\}_{j=0}^{n_0}, \{N_j\}_{j=0}^{n_1}, \{N_j\}_{j=0}^{n_2}, \ldots, \{N_j\}_{j=0}^{n_i}, \ldots$$

is thus the computation of every word in every nonrandom languages up to length 0, 1, 2, \ldots, $i$, \ldots. We conclude that every nonrandom language is RE, since, if language $L$ is nonrandom, $L^{[\leq k]}$ is computed by the Turing Machine which computes $\{N_j\}_{j=0}^{n_k}$, for all $k \in \mathbb{N}$.

If $L$ is recursively enumerable, the characteristic function of $L$, namely, $\chi_L : A^* \rightarrow \mathbb{N}$ such that

$$\chi_L(w) = \begin{cases} 1, & \text{if } w \in L \\ 0, & \text{if } w \notin L \end{cases}$$

is computable. Thus, the composition $\chi_L \circ l^{-1}$ is a computable function, and the sequence $N = \{n_i\}_{i \in \mathbb{N}}$, defined as follows, gives the recursively enumerable sequence for the randomness test of which $L$ is the only member. For each $i \in \mathbb{N}$, let the integers $s_i$, $t_i$, and $n_i$ be defined as follows. Let

$$s_i = \sum_{j=0}^{2^{\frac{n_i-1}{\alpha-1}}} 2^j [\chi_L \circ l^{-1}(j)],$$

i.e., the number $\tilde{l}(L^{[\leq k]})$. Let

$$t_i = i - \min \left\{ k : 2^{\frac{n_k+1}{\alpha-1}} > s_i \right\},$$

i.e., the difference between $i$ and the length of the longest word in $L^{[\leq i]}$. Let $n_i = \langle i, \langle s_i, t_i \rangle \rangle$, so that

$$E_{n_i} = C_{L^{[\leq i]}},$$

Define $\mathcal{V} = \{V_i\}_{i \in \mathbb{N}}$, where $V_i = \bigcup_{j \in \mathbb{N}} E_{n_j}$. Then, on the one hand, by construction $V_i$ is a single cylinder set of length $i$, i.e., of measure $2^{-\frac{n_i-1}{\alpha-1}} \ll 2^{-i}$. On the other hand, selecting any element of $V_i$, for each $i \in \mathbb{N}$, gives a language sequence which converges to $L_i$; that is, any language sequence
\[ L = \{L_k\}_{k \in \mathbb{N}} \text{ such that } L_k \in V_i \text{ converges to } L. \] Thus, the family RE and the set of nonrandom languages under the Cantor topology are equivalent.

Recall from Corollary 2.2.1 that the finite languages, which are certainly RE, are dense in the Cantor space. Therefore, from the second part of Theorem 1.3, the non-RE languages are meagre in \((\mathcal{P}_A, E, \mu)\), i.e., they are contained in a countable union of nowhere dense sets.

We move on to a different pseudo-metric altogether, one which takes the entire set-difference of two languages into account.
This chapter is a detailed exploration of a language pseudo-metric quite different from Cantor distance. These are some of the motivations for seeking out such an alternative: under the Cantor topology, the finite languages are dense in the language space; under the Cantor topology, a language space essentially is a copy of the Cantor space; the Cantor space itself is totally disconnected, so there is no hope for connectedness under the topology induced by the Cantor distance. We define a distance function on languages that is, in some respects, more satisfactory than the metric $d_1$. It is shown, for instance, that neither finite nor locally testable languages are dense in the topology induced by this distance. However, it is also shown that this is still not enough to make a distinction, on a topological basis, between regular and linear languages. Random languages are still non-RE in the randomness space determined by the topology of this new distance, yet, at the same time, uncountably many non-RE languages are found to be nonrandom.

We call this new pseudo-metric Besicovitch distance. Under the topology induced by Besicovitch distance, which we call the Besicovitch topology, a language space is not compact. It has a geometry which becomes apparent from the vantage point of a metric quotient space. A point in the quotient space is an equivalence class in the language space. As an example of the distinct geometry visible from the quotient space, every equivalence class has a unique antipode, a single equivalence class from which it is at a maximum distance. A second vantage point, an upper quotient space is shown to be homeomorphic to the unit interval. From an examination of the upper quotient space, other aspects of the Besicovitch topology become apparent. For example, the otherwise elusive regular languages clearly map to a dense set in the upper quotient space. Certain of these observations will motivate the incorporation into a third pseudo-metric, in the following chapter, of a well-known feature of languages which quantifies the information a language is capable of expressing.

We will denote the Besicovitch distance by $d_\zeta$. The distance between two languages is decided by their symmetric set difference, but not any specific word-length. Rather, $d_\zeta$ is the upper-density of
the set difference, meaning the least upper bound on limits of ratios between the cardinality of sections of the symmetric set-difference to the cardinality of sections of \( A^* \). The Besicovitch distance is the development, with application to languages, of an analog to the pseudo-metric introduced by Cattaneo et al., in [7], with application to the configuration space \( S^Z \) of one-dimensional cellular automata. Parallel to our aim of avoiding the denseness of finite languages in a language space, so the stated aim, in [7], [3], and [12], of avoiding the density of shifted translations of cellular automata configurations in the configuration space; just as we are seeking a metric distinct from Cantor distance, so in [7, 3, 12], the problem stems from a metric under which the configuration space becomes yet another copy of the Cantor space. In fact, on biinfinite sequences, the word metric mentioned in the preface to the last chapter is the metric standardly used on the space of cellular automata configurations. Under this metric, \( d_C \), two biinfinite sequences \( x \) and \( y \) are at distance 0 they are equal, but otherwise \( d_C (x, y) = 2^{-k} \), where \( k = \min \{|i| : x[i] \neq y[i]\} \).

In the first section of this chapter, the Besicovitch language distance \( d_\zeta \) and language norm \( \|\cdot\|_\zeta \) are defined. In the second section, the surjectivity of the norm is proven. In the third section, the quotient space induced by Besicovitch distance on the language space is exhibited and analyzed, non-compactness is demonstrated, and the upper quotient space is introduced. The geometries of the language space, the first quotient space, and the second are examined in the fourth section. In section five, properties of the two quotient spaces are used to locate the various families of the Chomsky hierarchy within a language space. Finally, in section six, an analysis is presented of random languages in this space. Also presented are the two results regarding the non-RE languages with respect to randomness in the Besicovitch topology.

Before beginning, though, it is appropriate to recall the historical origin of Besicovitch pseudo-metrics and how it was adapted for use with cellular automata. The original Besicovitch pseudo-metric was developed for the investigation of almost-periodic real-valued functions in [2]. The distance between two almost-periodic real valued functions \( \phi \) and \( \psi \) in \( \ell^1 \) is expressed by the pseudo-metric

\[
d_{B^p}(\phi, \psi) = \limsup_{n \to \infty} \frac{1}{2n + 1} \sum_{i = -n}^{n} |\phi(x) - \psi(x)|,
\]

i.e., by the upper density of the mean of the ordinary \( \ell^p \)-norm of the difference between the functions as the interval of evaluation widens symmetrically around \( x = 0 \). Because the Besicovitch pseudo-metric depends on the evaluation of the two functions only at discrete intervals, it is nat-
urally adaptable to expressing distances between objects with a bound proportion of differences, where differences can be evaluated around a central point at countably many intervals.

Cellular automata (CAs) are precisely such objects. The subject of lay speculation as well as deep research, the classification of one-dimensional CAs is still controversial. The descriptive classifications of [39] have given way to a battery of precisely-defined characteristics [3, 7, 12]. A one-dimensional cellular automaton \( \Phi \) is a discrete time dynamical system which iteratively assigns a symbol from an alphabet \( S \) to each coordinate of a biinfinite array according to a local rule \( \phi \). The cellular automaton \( \Phi \) is, therefore, a map from \( S^x \) into \( S^y \) on the basis of some frame-size \( k \in \mathbb{N} \) and some local function \( \phi \), which maps \( S^{(2k+1)} \) into \( S \). The space \( S^x \) is called the phase space or configuration space; an element \( x \) of \( S^x \) is called a configuration; and each coordinate, or cell \( x_{[i]} \), where \( i \in \mathbb{Z} \), of configuration \( x \in S^x \) may be identified with the symbol of \( S \) which it contains. That is, we can write \( x_{[i]} = s \in S \). Then \( \Phi \) maps configuration \( x \) into a configuration \( y \) by the global rule

\[
y_{[i]} = [\Phi \phi (x)]_{[i]} = \phi (x_{[j]}, i - k \leq j \leq i + k).
\]

The conception is that, if \( x \) is the configuration of the CA at time \( t \), then \( y \) is the configuration of the CA at time \( t + 1 \). Over an alphabet of two symbols \( \{0, 1\} \), a one-dimensional CA configuration is a biinfinite binary sequence. In [7], the Besicovitch pseudo-metric \( d_B \) on the space of one-dimensional configurations over the alphabet \( \{0, 1\} \) is defined as follows:

\[
d_B(x, y) = \limsup_{k \to \infty} \frac{\# \{ i : |i| \leq k \text{ and } x_{[i]} \neq y_{[i]} \}}{2k + 1}.
\]

Thus, \( d_B \) is the the upper mean of the Hamming distance between subsequences centered at position 0 in two biinfinite sequences.

The Besicovitch distance on a language space may be considered the generalization of \( d_B \) to an alphabet. That is, where \( d_B \) is a distance based upon the two possibilities at each cell, namely, it is either occupied by a 1 or a 0, we extend to a function operating on word lengths, where at every word length \( k \), the quantity \( \#L \triangle^k M \) is an integer between 0 and \( \alpha^k \). In a sense, therefore, \( d_B \) is the identical pseudo-metric, but over an alphabet consisting of one symbol, where only one-sided configurations are considered (i.e., in the space \( \{0, 1\}^N \)). Then the expression \( \# \{ i : |i| \leq k \text{ and } x_{[i]} \neq y_{[i]} \} \) gives the number of words in the set-difference of two languages \( L_x \) and \( L_y \) defined such that
$A = \{a\}$ and, for $z \in \{0,1\}^N$, the word $a^k \in L_y$ if and only if $z[k] = 1$. Then, replacing $2k + 1$ by $k - 1$, we have

$$d_B|_{\{0,1\}^n} (x, y) = \limsup_{k \to \infty} \frac{\# (L_x \triangle L)}{k}$$

The Besicovitch distance is a strict pseudo-metric, not a metric, for languages as for CAs and almost periodic functions. The first quotient space is also used in [7, 3]; the second quotient space, and the link to the unit interval are elaborated more explicitly here.

3.1 A Besicovitch pseudo-metric on language spaces

If a language is infinite, its cardinality, like the Cantor distance, hides more than it reveals about the size of the language. Another intuitive notion of the size of a language would involve consideration of the cardinality of each section of the language to the cardinality of all possible words up to that length. That is, we consider the ratio of the number of words in the language up to length $k$, $\#L[\leq k]$, to $\#A[\leq k]$. As the length of the section goes to infinity, the ratio mentioned a term in a sequence of ratios which is bound above by one and below by zero. Consequently, it is assured that, by inspecting the sequence of ratios, we can encapsulate size in a way that is possibly more meaningful than cardinality: we can give the least upper bound of limits of the ratio of sizes of sections of the language to the size of the corresponding section of $A^*$. We show in the next section that this notion of size is fundamentally different from the Cantor language norm.

**Definition 3.1.1** Let $\|\cdot\|_\zeta$ be the function $\|\cdot\|_\zeta : \mathcal{P}_A \to \mathbb{R}^\geq 0$, for fixed alphabet $A$, such that

$$\|L\|_\zeta = \limsup_{k \to \infty} \frac{\#L[\leq k]}{\#A[\leq k]}$$

(3.1)

Let $d_\zeta$ be the function $d_\zeta : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}^\geq 0$ such that $d_\zeta : (L, M) \mapsto \|L \triangle M\|_\zeta$, i.e.

$$d_\zeta (L, M) = \limsup_{k \to \infty} \frac{\# (L \triangle^c M)}{\#A[\leq k]}.$$  (3.2)

If language $L$ is a subset of language $M$, then, for each $k \in \mathbb{N}$, $\#L[\leq k] \leq \#M[\leq k]$. It follows that the function $\|\cdot\|_\zeta$ is a language norm, which we will call the Besicovitch language norm. Therefore,
by Proposition 1 on page 27, the function \( d_\zeta \) is a language pseudo-metric. We will call \( d_\zeta \) the Besicovitch distance.

The norm \( \| \cdot \|_\zeta \) takes the same value on language \( L \) as on the language \( \pi (L) \), where \( \pi \) is a permutation of the symbols of \( A \), since, if \( |w| = k, |\pi (w)| = k \) for all \( \pi \in S_\alpha \). Therefore \( \| \cdot \|_\zeta \) is a permutative language norm.

Besicovitch distance, however, may have a different value for different \( \pi \)-differences between languages, although the cardinality of each \( \pi \)-difference may be infinite. Thus, the permutative version of Besicovitch distance between two languages is the \( \pi \)-difference of minimum norm over the permutation group.

\[
d^{(\alpha)}_\zeta (L, M) = \min_{\pi \in S_\alpha} \limsup_{k \to \infty, k > 0} \frac{\# (L \triangle \pi M)}{\# A^k}.
\]

By definition, \( d^{(\alpha)}_\zeta (L, M) \leq d_\zeta (L, M) \) for all languages \( L, M \in \mathcal{P}_A \). Much of what follows applies both to Besicovitch distance and to its permutative version. We will make some note of facts which are not true of both. Beyond that, this is a subject for further research.

**Remark 6** The general features of Besicovitch distance on language spaces are as follows:

1. The range of Besicovitch distance and language norm on language spaces is the unit interval.
   In other words, the definition could say \( d_\zeta : \mathcal{P}_A \times \mathcal{P}_A \to [0, 1] \). This is, in fact, a continuous map. Surjectivity will be demonstrated in Lemma 3.2 and Corollary 3.2.1. Continuity will be shown as a consequence of Corollary 3.11.1.

2. The Besicovitch distance between a language and its complement is 1. That is, \( d_\zeta (L, L^c) = 1 \).
   This follows from the fact that \( (L \cap A^k) \triangle (L^c \cap A^k) = A^k \) for each \( k \in \mathbb{N} \).

3. On every language space \( \mathcal{P}_A \), the Besicovitch distance is a strict pseudo-metric; for consider the languages \( L = \Lambda \) and \( M = \{a\} \), where \( a \in A \). Then \( L \neq M \), yet \( \# (L \triangle \alpha^k M) = 1 \), for all \( k \in \mathbb{N} \), and so
   \[
d_\zeta (L, M) = \lim_{k \to \infty} \frac{1}{\# A^k} = 0.
\]
   The permutative version of Besicovitch distance is also a strict pseudo-metric.
4. By (1.1) on page 7, (3.2) may be written as follows:

\[
\begin{align*}
    d_\zeta (L, M) &= \limsup_{k \to \infty} \frac{\# (L \triangle^k M)}{(\frac{\alpha^k - 1}{\alpha - 1})} = \limsup_{k \to \infty} \left[ \frac{\# (L \triangle^k M)}{(\alpha^k - 1)} \right] \left( \frac{\alpha - 1}{\alpha^k - 1} \right), \\
    \end{align*}
\]

The permutative version of the distance may likewise be written

\[
\begin{align*}
    d_\zeta (L, M) &= \limsup_{k \to \infty} \frac{\# (L \pi \triangle^k M)}{(\frac{\alpha^k - 1}{\alpha - 1})} = \min_{\pi \in S_\alpha} \limsup_{k \to \infty} \left[ \frac{\# (L \pi \triangle^k M)}{(\alpha^k - 1)} \right] \left( \frac{\alpha - 1}{\alpha^k - 1} \right). \\
    \end{align*}
\]

If languages \( L \) and \( M \) are disjoint, then the cardinality of words of any length \( k \) in the union of the two languages is equal to the sum \( \#L[k] + \#M[k] - \#(L \cap M)[k] \), which by assumption is just \( \#L[k] + \#M[k] \). This proves the following lemma.

**Lemma 3.1** If \( L \) and \( M \) are disjoint languages in \( \mathcal{P}_A \), then \( \|L\|_\zeta + \|M\|_\zeta = \|L \cup M\|_\zeta \).

The necessary condition for additivity of the norm follows from this.

**Corollary 3.1.1** If \( L \) and \( M \) are languages in \( \mathcal{P}_A \), then \( \|L\|_\zeta + \|M\|_\zeta = \|L \cup M\|_\zeta \) if and only if \( \|L \cap M\|_\zeta = 0 \).

**Proof.** By Lemma 3.1, \( \|L\|_\zeta + \|M\|_\zeta = \|L\|_\zeta + \|M\setminus L\|_\zeta + \|L \cap M\|_\zeta = \|L \cup M\|_\zeta + \|L \cap M\|_\zeta \), which is equal to \( \|L \cup M\|_\zeta \) if and only if \( \|L \cap M\|_\zeta = 0 \). \( \square \)

But it follows from the lemma also that the norm is truly “norm-like”. For the addition operation on the semiring \( \mathcal{P}_A \) is set union, and we have the following.

**Corollary 3.1.2** For all languages \( L \) and \( M \) in \( \mathcal{P}_A \), \( \|L\|_\zeta + \|M\|_\zeta \geq \|L \cup M\|_\zeta \).

**Proof.** By Lemma 3.1,

\[
\begin{align*}
    \|L\|_\zeta + \|M\|_\zeta &= \|L\|_\zeta + \|L \cap M\|_\zeta + \|M\setminus L\|_\zeta = \|L \cup M\|_\zeta + \|L \cap M\|_\zeta \geq \|L \cup M\|_\zeta. \\
    \end{align*}
\]

\( \square \)
We can immediately contrast the norm $\|\cdot\|_\zeta$ with the norm $\|\cdot\|_1$. The Cantor norm maps every language to a power of $\frac{1}{2}$. We proceed to show that the range of the Besicovitch language norm is the entire unit interval.

### 3.2 The Besicovitch language norm is surjective

We claim that $\|\mathcal{P}_A\|_\zeta = [0, 1]$, i.e., that the image of the entire language space under the Besicovitch norm is precisely the entire the unit interval. To justify this claim, we need to be able to construct a language with an arbitrary norm. We identify a collection of such languages.

**Definition 3.2.1** Given $r \in [0, 1]$ and $\alpha \in \mathbb{N}$ such that $\alpha > 1$, let $\tilde{r}_{(\alpha)}$ denote the sequence $\{\tilde{r}_k\}_{k \in \mathbb{N}}$ where $\tilde{r}_k = \lfloor r \alpha^k \rfloor$. Let the set of $r$-simple languages in $\mathcal{P}_A$, denoted $L_r$, be the set of languages $L$ in $\mathcal{P}_A$ such that $\#L[i] = \tilde{r}_i$ for all $i \in \mathbb{N}$. In other words,

$$L_r = \left\{ L \in \mathcal{P}_A : \left\{ \#L[i] \right\}_{i \in \mathbb{N}} = \tilde{r}_{(\alpha)} \right\}.$$

Notice that it is possible to construct an $r$-simple language for any $r \in [0, 1]$ because, for each $k \in \mathbb{N}$, $\tilde{r}_k < \alpha^k$. The last sentence of the definition simply states that the selection of a set of $\tilde{r}_k$ words from $A^k$ for all $k \in \mathbb{N}$ results in a sequence of languages, the cardinalities of which are equivalent to the sequence $\{\tilde{r}_k\}_{k \in \mathbb{N}}$.

**Lemma 3.2** If $r \in [0, 1]$, there is at least one $r$-simple language; every simply $r$ language has norm $r$.

**Proof.** We claim, in other words, that there is a language with a norm of any value $r$ in the unit interval. By construction for each $r \in [0, 1]$ the sequence $\tilde{r}_{(\alpha)}$ exists. Since $0 \leq \tilde{r}_k < \alpha^k$, there are at least $\tilde{r}_k$ words in $A^k$ for all $k \in \mathbb{N}$. Therefore, the set $L_r$ is non-empty for each $r \in [0, 1]$. If $L$ is a language in the set $L_r$, then by Definition 3.1.1,

$$\|L\|_\zeta = \limsup_{k \to \infty} \frac{\sum_{i=0}^{k-1} \tilde{r}_i}{\sum_{i=0}^{k-1} \alpha^i}.$$

56
From the way in which \( \hat{r} \) was constructed,

\[
\|L\|_\zeta \leq \limsup_{k \to \infty} \frac{\sum_{i=0}^{k-1} \hat{r}_i}{\sum_{i=0}^{k-1} \alpha^i} \leq \frac{\sum_{i=0}^{k-1} r a^i}{\sum_{i=0}^{k-1} \alpha^i} = r \leq \limsup_{k \to \infty} \frac{\sum_{i=0}^{k-1} (\hat{r}_i + 1)}{\sum_{i=0}^{k-1} \alpha^i} \\
\leq \|L\|_\zeta + \lim_{k \to \infty} \frac{k(\alpha - 1)}{\alpha^k - 1}
\]

(3.3)

But then, of course, \( \|L\|_\zeta = r \), which establishes the claim. \( \Box \)

This establishes the claim stated at the beginning of this section. We restate it, for clarity.

**Corollary 3.21** The Besicovitch language norm \( \| \cdot \|_\zeta \) is a surjective mapping from \( \mathcal{P}_A \) onto \( [0, 1] \).

The \( r \)-simple languages are a large, in fact uncountable set for each \( r \in [0, 1] \). They also contain interesting subsets. Consider, for example, the following special type of ideal over \( A^* \).

**Definition 3.22** We will call \( I \) a right, left, or two-sided word ideal, respectively, of the monoid \( A^* \) if there is a word \( w \in A^* \) such that \( I = wA^* \), \( I = A^*w \), or \( I = A^*wA^* \), respectively.

If \( r = 0 \), then there is no ideal of \( A^* \) in \( L_0 \), for reasons shown later (Corollary 3.12.1).

**Lemma 3.3** If the real number \( r \) is such that \( r \in [0, 1] \), there exists a right ideal of \( A^* \) in \( L_r \).

**Proof.** If \( r = 1 \), then \( w = \lambda \) trivially satisfies the claim of the lemma. Suppose therefore that \( r \in (0, 1) \). Since, by definition, \( 0 \leq \hat{r}_1 < \alpha \), there is a subset \( I_1 \) of \( A \) (actually, at least \( \alpha \) subsets) such that \( \#I_1 = \hat{r}_1 \). Note from the definition of \( L_r \),

\[
\hat{r}_k \leq r\alpha^k < \hat{r}_k + 1
\]

(3.4)

for all \( k \in \mathbb{N} \). Multiplying through by \( \alpha \) gives the inequality

\[
\hat{r}_k \alpha \leq r\alpha^{k+1} < \hat{r}_k \alpha + \alpha.
\]

(3.5)

But for \( k + 1 \), we have, from

\[
\hat{r}_{k+1} = \left\lfloor r\alpha^{k+1} \right\rfloor \leq r\alpha^{k+1} < \hat{r}_{k+1} + 1.
\]

(3.6)
Since all values are non-negative integers, combining (3.5) and (3.6), this means

\[ \tilde{r}_k \alpha \leq \tilde{r}_{k+1} < \tilde{r}_k \alpha + \alpha. \]  

(3.7)

It follows that

\[ \tilde{r}_{k+1} = \tilde{r}_k \alpha + t_k, \]  

(3.8)

and that, for some \( t_k \in \mathbb{N} \) such that \( 0 \leq t_k < \alpha \), for all \( k \in \mathbb{N} \). Therefore, for all \( k \in \mathbb{N} \),

\[ \tilde{r}_k \alpha \leq \alpha^{k+1} - \alpha. \]  

(3.9)

Thus, there exists language \( T_1 \subseteq A^2 \setminus I_1 A \) such that \( \#T_1 = t_1 \), so that \( \#(I_1 A \cup T_1) = \tilde{r}_2 \). Set \( I_2 = I_1 A \cup T_1 \). Continuing in this fashion, let \( T_k \), for each \( k \in \mathbb{N} \), be a language such that \( T_k \subseteq A^{k+1} \setminus I_k A \) and \( \#T_k = t_k \). Finally, for \( k \in \mathbb{N} \) define language \( I \in \mathcal{P}_A \) such that \( I^{[k]} = I_k \), which is to say, let \( I \) be the union \( \bigcup_{i \in \mathbb{N}} I_i \). Then \( I \), by construction, is an element of \( L_r \), and \( w A^j \subseteq I \) for all \( w \in I \) and every \( j \in \mathbb{N} \). Thus, \( IA^* \subseteq I \). □

We will use a family of two-sided word ideals later. It is enough to note that every such ideal is a regular language, so we already know that regular languages are widely distributed in the language space.

### 3.3 Besicovitch distance quotient space

In order to proceed further, we need to use the property of convergence. We discuss first a quotient of a language space in which a point is an equivalence class and the equivalence relation identifies all languages at distance 0 from each other. The implied quotient map sends the pseudo-metric topology on the language space to a metric topology on the quotient space. In the first two subsections, we define distance on the Besicovitch quotient space and show that the quotient map is an isometry. In subsection three, we discuss the notion of convergence under the quotient metric and show the existence of an underlying sequence of indices expressing the convergence. In subsection four, we draw conclusions about the basic topological character of the language space under the Besicovitch distance; for instance, Lemma 3.7 shows that the Besicovitch topology is not compact. In the fifth subsection, we define an equivalence relation on points in the Besicovitch quotient space,
and show that this implies the homeomorphism of an upper quotient space with the unit interval. The latter enables us to discuss the geometry of a language space and language family distribution within it under the Besicovitch topology.

3.3.1 The Besicovitch distance equivalence relation and induced quotient space

This construction of a quotient space is also used in [7] with respect to CA configurations, and is well-known in analysis. We map each collection of languages at distance zero from each other to a point in a quotient space. The quotient space can then be metrized.

**Definition 3.3.1** Given \(L, M \in \mathcal{P}_A\), let \(L \sim_\zeta M\) if \(d_\zeta(L, M) = 0\), i.e., if \(\|L \Delta M\|_\zeta = 0\).

**Fact 7** The relation \(\sim_\zeta\) is an equivalence relation on \(\mathcal{P}_A\).

**Proof.** Reflexivity and symmetry are obvious. Let \(L, M, N \in \mathcal{P}_A\) such that \(L \sim_\zeta M\) and \(M \sim_\zeta N\). Since \(d_\zeta\) is a pseudo-metric, \(0 = d_\zeta(L, M) + d_\zeta(M, N) \geq d_\zeta(L, N)\). From Remark 6(1), \(L \sim_\zeta N\). □

**Definition 3.3.2** Given language space \(\mathcal{P}_A\),

1. The \(\sim_\zeta\) equivalence class of language \(L \in \mathcal{P}_A\) will be denoted \([L]_\zeta\).

2. The collection of \(\sim_\zeta\) equivalence classes will be called the Besicovitch quotient space over \(\mathcal{P}_A\), which we will denote \(\mathcal{Q}_A^\zeta\). Elements of the quotient space, i.e., points in \(\mathcal{Q}_A^\zeta\), will be denoted with sans serif letters, \(L, M, N, \ldots\). Collections of points will be denoted in bold-face sans serif, \(L, M, N, \ldots\).

3. Let \(\eta_\zeta\) denote the quotient mapping \(\eta_\zeta : \mathcal{P}_A \rightarrow \mathcal{Q}_A^\zeta\) which takes a language in \(\mathcal{P}_A\) to its \(\sim_\zeta\) equivalence class in \(\mathcal{Q}_A^\zeta\).

Since \(\mathcal{Q}_A^\zeta\) is a partition of \(\mathcal{P}_A\), the mapping \(\eta_\zeta\) is well-defined and surjective on \(\mathcal{Q}_A^\zeta\), but not injective since a single point of \(\mathcal{Q}_A^\zeta\) is the image under \(\eta_\zeta\) of every language in a \(\sim_\zeta\) equivalence class.

Since \(\mathcal{P}_A\) is a semiring under set union, the set operations of union, intersection, and complementation are preserved by mappings from collections of points in \(\mathcal{Q}_A^\zeta\) to the sets of languages of which they are the equivalence classes. In particular, every topology on \(\mathcal{Q}_A^\zeta\) is the quotient of a topology on \(\mathcal{P}_A\).
When the language $L \in \mathcal{P}_A$ is a member of the language family $L \subseteq \mathcal{P}_A$, and every member of $L$ is contained in one of the equivalence classes in the collection of points $L \subseteq Q_A$, then we can write $L \in \zeta L$ and $L \subseteq \zeta L$ in place of the clumsy $\eta_\zeta (L) = L$ and $\eta_\zeta (L) \subseteq L$. Since $\eta_\zeta$ is not invertible, the relationships $\in \zeta$ and $\subseteq \zeta$ state the location of a language or collection of languages in the quotient space and are not descriptions of the quotient space itself. Since the distinction between languages and collections of languages is indicated by the notation we have chosen, the subscript $\zeta$ will be dropped where we discuss the location of languages and sets of languages in the quotient space (so long as the pseudo-metric is understood).

The relation of $\sim_\zeta$-equivalence, by definition, is decided by the limit supremum of a sequence of ratios. But $\sim_\zeta$-equivalence, in turn, implies an upper bound on the the cardinality of sections of the symmetric set-difference of languages.

**Lemma 3.4** If languages $L, M \in \mathcal{P}_A$, then $L \not\sim_\zeta M$ if and only if there exists $m \in \mathbb{N}$ such that, for all $N \in \mathbb{N}$, there exists $k > N$ such that $\# (L \triangle^k M) \geq \# A^{k-m}$.

**Proof.** ($\Rightarrow$) We claim that if, for any $k' \in \mathbb{N}$, there is a $k > k'$ such that $\# (L \triangle^k M) < \# A^{k-m}$ for all $m \in \mathbb{N}$, then the Besicovitch distance between the two languages is 0. Of course $\# A^k \geq \alpha^{k-1}$. Let $k_0 = 0$ and let $k_m$ be the least integer greater than $k_{m-1}$ such that $\# (L \triangle^k M) < \# A^{k-m}$, for all $k > k_m$ for $m > 0$. The sequence $\{k_m\}_{m \in \mathbb{N}}$ is non-decreasing by assumption and our construction, so, if $k > k_m$,

\[
\frac{\# (L \triangle^k M)}{\# A^k} < \frac{\alpha^{k-m}}{\# A^k} < \frac{\alpha^{k-1} \alpha^{-m}}{\# A^k} = \alpha^{-m}.
\]

Taking the limit as $m$ goes to infinity, this implies

\[
\lim_{k \to \infty} \frac{\# (L \triangle^k M)}{\# A^k} = 0.
\]

($\Leftarrow$) We claim that the condition $\# (L \triangle^k M) \geq \# A^{k-m} = \alpha^{k-m}$ for some $k > N$ for all $N \in \mathbb{N}$ implies $L \not\sim_\zeta M$. We observe that $\# A^k < \alpha^k$. This implies the existence of a sequence of integers $\{k_i\}_{i \in \mathbb{N}}$ such that, for all $i \in \mathbb{N}$,

\[
\frac{\# (L \triangle^{k_i} M)}{\# A^{k_i}} \geq \frac{\alpha^{k_i-m}}{\# A^{k_i}} = \frac{\alpha^{k_i} \alpha^{-m}}{\# A^{k_i}} \geq \alpha^{-m}.
\]
But then \( \alpha^{-m}/2 \) is less than the upper bound of the limits of sequences of the form

\[
\left\{ \frac{\# \left( L \triangle^{<k_i} M \right)}{\# A^{<k_i}} \right\}_{i \in \mathbb{N}}
\]

where \( k_i \in \mathbb{N} \) for all \( i \), and hence

\[
d_\zeta (L, M) > \alpha^{-m}/2 > 0,
\]

i.e., \( d_\zeta (L, M) \neq 0 \), and hence \( L \not\sim_\zeta M \).

This way of looking at \( \sim_\zeta \)-equivalence points out the existence of special sequence of word-lengths for any two languages.

**Definition 3.3.3** Given languages \( L, M \in \mathcal{P}_A \), we will denote by \( K_\zeta (L, M) \) the integer sequence \( \{k_m\}_{m \in \mathbb{N}}, N \subseteq \mathbb{N} \). If \( m \in N \), then \( k_m > m \) and, for sections of \( L \triangle M \) longer than \( k_m \), the estimate of the distance of the two languages is bound above by \( \alpha^{-m} \), i.e., such that, if \( k > k_m (> m) \), then

\[
\frac{\# \left( L \triangle^{<k} M \right)}{A^{<k}} < \alpha^{-m}
\]

and so \( \# \left( L \triangle^{<k} M \right) < \# A^k = \alpha^k \).

The sequence \( K_\zeta (L, M) \) expresses concisely the relative location of languages in the quotient space \( Q^A_\zeta \), i.e., the cardinalities of sections of the symmetric set difference of languages \( L \) and \( M \). This will be developed further in subsection 3.4.1.

3.3.2 The metric quotient topology

We define the metric \( d_\zeta \) on the Besicovitch quotient space \( Q^A_\zeta \) of \( \mathcal{P}_A \) as the lifting of Besicovitch distance to the quotient space. It becomes apparent that the quotient map \( \eta_\zeta \) is an isometry.

**Definition 3.3.4** Let the distance \( d_\zeta \) between points \( L \) and \( M \) in \( \mathcal{Q}^A_\zeta \) be defined as

\[
d_\zeta (L, M) = \inf \left\{ d_\zeta (L, M) : L \in \eta^{-1}_\zeta (L), M \in \eta^{-1}_\zeta (M) \right\}
\]

(3.10)

**Lemma 3.5** \( L, M \in \mathcal{Q}^A_\zeta \), then \( d_\zeta (L, M) = 0 \) if and only if \( L = M \).
Proof. (⇐) Trivial.

(⇒) The claim is that, if \(d_\zeta (L, M) = 0\), then \(L = M\). Suppose language \(L \in \mathcal{P}_\zeta\) is such that \(L \in L\), but \(L \notin M\). We conclude by Definition 3.3.4 that there exists \(M \in M\), such that \(d_\zeta (L, M) = \varepsilon > 0\). But then, for arbitrary languages \(L' \in L\) and \(M' \in M\), we have by the triangle inequality that

\[
\varepsilon \leq d_\zeta (L, M) \leq d_\zeta (L, L') + d_\zeta (L', M') + d_\zeta (M', M) \leq d_\zeta (L', M').
\]  

(3.11)

Thus \(d_\zeta (L, M) \geq \varepsilon/2 > 0\), by Definition 3.3.4, contrary to our hypothesis. \(\Box\)

Now we show that the quotient map \(\eta_\zeta\) is an isometry. When we talk about Besicovitch distance between languages, from now on, we can, interchangeably discuss distance between points in \(\Omega_\zeta^A\).

COROLLARY 3.5.1 \(L, M \in \mathcal{P}_\zeta\), the diagram in Figure 1 commutes.

In other words,

\[
d_\zeta (\eta_\zeta (L), \eta_\zeta (M)) = d_\zeta (L, M).
\]

Proof. If \(L \sim_\zeta M\), then \(\eta_\zeta (L) = \eta_\zeta (M)\), and we are done, by Lemma 3.5. But conversely, assume that \(L \not\sim_\zeta M\), with \(\eta_\zeta (L) = L\) and \(\eta_\zeta (M) = M\). Then \(0 < d_\zeta (L, M) \leq d_\zeta (L, M)\) by Definition 3.3.4. Suppose, toward contradiction, that \(d_\zeta (L, M) < d_\zeta (L, M)\). Then \(d_\zeta (L, M) =\)
\( \text{d}_\xi(L, M) + \varepsilon, \) for some \( \varepsilon > 0, \) and there exist languages \( L' \in L \) and \( M' \in L \) such that

\[
0 < \text{d}_\xi(L, M) \leq \text{d}_\xi(L', M') < \text{d}_\xi(L, M) + \varepsilon/2 = \text{d}_\xi(L, M) - \varepsilon/2
\]

and so \( \text{d}_\xi(L, M) - \text{d}_\xi(L', M') > \varepsilon/2. \) By the triangle inequality,

\[
\text{d}_\xi(L, L') + \text{d}_\xi(L', M') + \text{d}_\xi(M, M') \geq \text{d}_\xi(L, M)
\]

which implies that

\[
\text{d}_\xi(L, L') + \text{d}_\xi(M, M') > \varepsilon/2 > 0,
\]

so that either \( \text{d}_\xi(L, L') > 0 \) or \( \text{d}_\xi(M, M') > 0, \) meaning that either \( L \not\sim_\xi L' \) or \( M \not\sim_\xi M' \neq M, \) contrary to supposition. We conclude that \( \text{d}_\xi(L, M) = \text{d}_\xi(L, M) \) if \( \eta_\xi(L) = L \) and \( \eta_\xi(M) = M. \)

\[\square\]

**Corollary 3.5.2** The Besicovitch quotient space is a metric space under distance \( \text{d}_\xi. \)

**Proof.** Trivial. \[\square\]

**Corollary 3.5.3** If languages \( L \) and \( M \) are in point \( \Lambda \) of the Besicovitch quotient space, then

\[
\|L\|_\xi = \|M\|_\xi.
\]

**Proof.** The empty language \( \Lambda \in M, \) where \( M \) is a point of \( \Omega^A_\xi, \) by definition. But

\[
\|L\|_\xi = \text{d}_\xi(L, \Lambda) = \text{d}_\xi(L, M) = \text{d}_\xi(M, \Lambda) = \|M\|_\xi
\]

by Corollary 3.5.1. \[\square\]

The \( \text{d}_\xi \) metric topology on the quotient space is, therefore, the quotient of the pseudo-metric topology induced by Besicovitch distance on the language space.

**Definition 3.3.5** Let \( \tilde{\tau}_\xi \) denote the collection of open sets in \( \Omega^A_\xi \) under the \( \text{d}_\xi \) metric topology, and let \( \tau_\xi \) denote the collection of language sets in \( \mathcal{P}_A \) such that \( \eta(\tau_\xi) = \tilde{\tau}_\xi. \)
From Corollary 3.5.1, \((\mathcal{P}_A, \tau_\zeta)\) denotes the pseudo-metric topological space induced on \(\mathcal{P}_A\) by \(d_\zeta\).
We will call \(\tau_\zeta\) the Besicovitch (language) topology.

### 3.3.3 Convergence in the quotient space

Since, by Remark 6, the Besicovitch language topology is not \(T_1\), convergence to a language is not well-defined in \((\mathcal{P}_A, \tau_\zeta)\). There is no such difficulty in the quotient space.

**Lemma 3.6** A sequence of points \(L = \{L_i\}_{i \in \mathbb{N}}\) in \(\Omega^A_\zeta\) converges to the point \(L \in \Omega^A_\zeta\) if and only if, for all \(m \in \mathbb{N}\), there exists \(k_m \in \mathbb{N}\) such that \(i > k_m\) implies that if language \(L_i \in L_i\) and language \(L \in L\), then there exists integer \(N_i\), dependent only on \(i\), such that \(k > N_i\) implies \(#(L \triangle L_i) < \alpha^{k-m}\).

**Proof.** Follows directly from Lemma 3.4 and Corollary 3.5.1. \(\square\)

Note in particular that convergence does not require that the total distinctions between languages in a sequence and the limiting language go to zero as the number of the term of the sequence goes to infinity, as in the Cantor space. Instead, in the Besicovitch language topology, a convergent sequence of points converges to \(\sim_\zeta\) equivalence with languages in the limiting point. We can extend the definition of the sequence used in Lemma 3.12 to a sequence \(K_\zeta (L, L) = \{k_m\}_{m \in \mathbb{N}}\) of indices such that, if \(k > k_m\), then, if \(L_k \rightarrow L\), then, for every \(L_k\) in point \(L_k\) and every language \(L\) in \(L\), the cardinality of the symmetric set difference section \(L_k \triangle L\) exceeds the number of possible words of length \(l - m\) for at most finitely many \(l \in \mathbb{N}\). Thus the proportion of the cardinality of words in a sufficiently long section of the set-difference to the same-length section of \(A^*\) is bound above by \(\alpha^{-m}\).

### 3.3.4 The quotient space is perfect, but not compact

A primary concern for any topological space is whether the space is compact. In the Besicovitch quotient space, since it is a metric space, we can address this by determining whether every infinite sequence of points has a convergent subsequence. We show that neither the quotient space \(\Omega^A_\zeta\) nor the language space itself is compact, although \(\Omega^A_\zeta\) is a perfect set. This means the consideration of dynamical systems, i.e., iterated mappings, in this space will not be straightforward.
We proceed by appending three more corollaries to Lemma 3.6. These establish that the Besicovitch quotient space is perfect (i.e., that every point in the space is an accumulation point).

**Corollary 3.6.1** \( \eta_\zeta (\Lambda) \) is an accumulation point in \( (\Omega^A_\zeta, \tau_\zeta) \).

**Proof.** Let \( \{L_i\}_{i \in \mathbb{N}} \subseteq \mathcal{P}_A \) be a sequence of languages such that \( \|L_i\|_\zeta = \alpha^{-i} \) for all \( i \in \mathbb{N} \). Let \( \{L_i\}_{i \in \mathbb{N}} \) be the sequence of points \( L = \{L_i\}_{i \in \mathbb{N}} \) in \( \Omega^A_\zeta \) such that \( L_i \in L \) for all \( i \in \mathbb{N} \). This is possible by Corollary 3.2.1 and because \( \Omega^A_\zeta \) is a partition of \( \mathcal{P}_A \). Then \( L_i \to \eta_\zeta (\Lambda) \).

**Corollary 3.6.2** \( \eta_\zeta (A^*) \) is an accumulation point in \( (\Omega^A_\zeta, \tau_\zeta) \).

**Proof.** Consider a sequence of languages \( \{L_i\}_{i \in \mathbb{N}} \subseteq \mathcal{P}_A \) such that \( \|L_i\|_\zeta = 1 - \alpha^{-i} \).

**Corollary 3.6.3** \( (\Omega^A_\zeta, \tau_\zeta) \) is perfect.

**Proof.** We show that every point in \( \Omega^A_\zeta \) is an accumulation point. Suppose \( L \in L \), and \( L \in \Omega^A_\zeta \). If \( \|L\|_\zeta = 0 \) or \( \|L\|_\zeta = 1 \), then we are done, by Corollary 3.6.1 or Corollary 3.6.2. Suppose \( 0 < \|L\|_\zeta < 1 \). The claim is that there is a sequence of points in \( \Omega^A_\zeta \), not including \( L \), which converges to \( L \). Let point sequence \( \{M_i\}_{i \in \mathbb{N}} \), as in the proof of Corollary 3.6.1, be such that \( n \) which the terms are the equivalence classes of a sequence of languages \( \{M_i\}_{i \in \mathbb{N}} \) such that \( \|M_i\|_\zeta = \alpha^{-i} \) if language \( M_i \in M_i \). Then \( M_i \to \eta_\zeta (\Lambda) \). Construct a point sequence \( \{N_i\}_{i \in \mathbb{N}} \subseteq \Omega^A_\zeta \) as follows: let \( N_i = M_i \triangle \Lambda \) for some \( M_i \in M_i \) and each \( i \in \mathbb{N} \). Then, let \( N_i = \eta_\zeta (N_i) \).

We have \( \|N_i\|_\zeta = \|M_i\|_\zeta \) for all \( i \in \mathbb{N} \). Indeed, \( w \in N_i \triangle \Lambda \) if and only if \( w \in (M_i \setminus \Lambda) \setminus L = M_i \setminus L \), \( w \in (L \setminus M_i) \setminus \Lambda = \lambda \), or \( w \in L \setminus (M_i \setminus L) = L \cap M_i \), i.e., either \( w \in M_i \setminus L \) or \( w \in M_i \cap L \), which is to say, \( N_i \triangle \Lambda = (M_i \setminus L) \cup (M_i \cap L) = M_i \). So \( N_i \in \Omega^A_\zeta \setminus L \), because \( \|M_i\|_\zeta = \alpha^{-i} > 0 \) for all \( i \in \mathbb{N} \); yet \( N_i \to L \).

With regards to the permutative version of Besicovitch distance, the condition of \( r \)-simplicity, by itself, does not suffice for the construction of such convergent sequences as are used in the proof of Corollary 3.6.3. Therefore, surjectivity of the norm may not be enough to prove \( \mathcal{P}_A \) perfect under the permutative version of Besicovitch distance.

We now employ a family of two-sided word ideals in \( A^* \) which, when split into non-disjoint right ideals, yields an infinite sequence of points in the quotient space with no convergent subsequence. It
is yet to be determined whether a collection of sufficiently non-permutative languages can be found which will give the same result in the permutative case. That result is that neither \((Q_A, \bar{\tau}_\zeta)\) nor \((\mathcal{S}_A, \tau_\zeta)\) is compact.

**Definition 3.3.6** If \(w \in A^*\), let \(J_w\) denote the two-sided word ideal \(A^*wA^*\). Then, for \(k \in \mathbb{N}\), the \(k^{th}\) section of \(J_w\) is the right ideal \(A^k w A^* \subseteq J_w\), which will be denoted \(J_{w,k}\).

In regards to the \(J_w\) ideals and their sections, note that, while \(\bigcup_{k \in \mathbb{N}} J_{w,k} = J_w\), the intersection \(J_{w,i} \cap J_{w,j}\) is never empty. If \(l = |w|\), we can suppose that \(0 \leq i \leq j - l\), i.e., that the defining occurrences of the word \(w\) as a factor of words in \(J_{w,i}\) and \(J_{w,j}\) do not overlap (overlap occurs in at most finitely many cases). Then each word in \(J_{w,i}\) is of the form \(uwv\) where \(u \in A^i\), \(v \in A^{*}\), and each word in \(J_{w,j}\) is of the form \(u_1u_2u_3wv'\), where \(u_1 \in A^l\), \(u_2 \in A^l\), \(u_3 \in A^{j-i-l}\), and \(v' \in A^*\). Let \(B = A^l \setminus \{w\}\). By examination, \(J_{w,i} \setminus J_{w,j}\) is the set of words \(A^i w A^{<j+l-i}\) \(\cup A^i w A^{j-i-l} BA^*\), and \(J_{w,j} \setminus J_{w,i}\) is the set of word \(A^i BA^{j-i-l} w A^*\). Thus, for word-lengths \(k > j + l\), we have

\[
\#(J_{w,i} \setminus J_{w,j})^{<k} = \#\left(A^i w A^{<j+l-i}\right) + \#\left(A^i w A^{j-i-l} BA^{<k-j}\right),
\]

\[
\#(J_{w,j} \setminus J_{w,i})^{<k} = \#\left(A^i BA^{j-i+l} w A^{<k-j}\right),
\]

which can be expanded to give

\[
\#(J_{w,i} \setminus J_{w,j})^{<k} = \sum_{s=0}^{j+l-i-1} \alpha^{i+s} + \sum_{t=0}^{k-j-1} \alpha^{j-l+t},
\]

\[
\#(J_{w,j} \setminus J_{w,i})^{<k} = \sum_{u=0}^{k-j-1} \alpha^{j-l+u},
\]

and thence to

\[
\#(J_{w,i} \setminus J_{w,j})^{<k} = \alpha^i \left(\frac{\alpha^{j+l-i-1}}{\alpha - 1}\right) + \alpha^{j-l} \left(\frac{\alpha^{k-j-1}}{\alpha - 1}\right) = \frac{\alpha^{k-l} + \alpha^{j+l} - \alpha^{j-l} - \alpha^i}{\alpha - 1},
\]

\[
\#(J_{w,j} \setminus J_{w,i})^{<k} = \alpha^{j-l} \frac{\alpha^{k-j-1}}{\alpha - 1} = \frac{\alpha^{k-l} - \alpha^{j-l}}{\alpha - 1},
\]
which sum to give

\[
\# (J_{w,i} \triangle^k J_{w,j}) = \frac{2\alpha^{k-l} + \alpha^{j+l} - 2\alpha^{j-l} - \alpha^i}{\alpha - 1} = 2\alpha^{-l} \left[ \left( \frac{\alpha^k - 1}{\alpha - 1} \right) + \left( \frac{\alpha^{j+l}/2 - \alpha^j - \alpha^{i+l}/2 + 1}{\alpha - 1} \right) \right].
\]

Thus, the distance between \(J_{w,i}\) and \(J_{w,j}\) can be computed, as well as the norm of \(J_{w,i}\) (for any \(i \in \mathbb{N}\)), where \(|w| = l\).

\[
d_\zeta (J_{w,i}, J_{w,j}) = \limsup_{k \to \infty} 2\alpha^{-l} + \frac{\alpha^{j+l}/2 - \alpha^{j-l} - \alpha^{i+l}/2 + 1}{\alpha l (\alpha^k - 1)} = 2\alpha^{-l}.
\] (3.12)

\[
\|J_{w,i}\|_\zeta = \limsup_{k \to \infty} \frac{\# (A^i w A^{<k-i-l})}{\# A^{<k}} = \limsup_{k \to \infty} \frac{\sum_{s=0}^{k-i-l-1} \alpha^{i+s}}{\# A^{<k}} = \limsup_{k \to \infty} \alpha^{-l} \left[ \frac{\alpha^k - 1}{\alpha - 1} - \frac{\alpha^{i+l} - 1}{\alpha^k - 1} \right] = \alpha^{-l}.
\] (3.13)

From these calculations, setting \(J_{w,i} = \eta_{\zeta} (J_{w,i})\), the sequence \(\{J_{w,i}\}_{i \in \mathbb{N}}\) has the following properties: no subsequence of this sequence can converge, yet every language in each point of the sequence has the same norm.

**Lemma 3.7** The Besicovitch quotient space \((\mathcal{Q}_A^A, \tau_\zeta)\) is not compact.

**Proof.** It is sufficient to display an infinite sequence of languages belonging to distinct \(\sim_\zeta\) equivalence classes separated from each by a distance greater than some fixed \(\varepsilon\) such that \(\varepsilon > 0\). The idea is that the \(\eta_{\zeta}\) images of these languages will form an infinite sequence in \(\mathcal{Q}_A^A\) which has no convergent subsequence.

Consider the language sequence \(J_a = \{J_{a,i}\}_{i \in \mathbb{N}}\) where \(a \in A\). Two distinct terms \(J_{a,i}\) and \(J_{a,j}\) are at distance \(2\alpha^{-1}\), from (3.12), so consider the point sequence \(L = \{L_i\}_{i \in \mathbb{N}}\), where \(J_{a,i} \in L_i\) for
all \( i \in \mathbb{N} \). By Corollary 3.5.1, there is no convergent subsequence of \( L \), since \( d_\zeta (L_i, L_j) > \alpha^{-1} \) if \( i \neq j \).

Sequential compactness is not defined in a pseudo-metric space, so we add the following corollary to clear up any doubt about compactness in the language space.

**Corollary 3.7.1** The metric \( d_\zeta \) is not complete.

**Proof.** It suffices to exhibit a sequence of points which are Cauchy convergent in \( Q^A_\zeta \), but which do not converge to any point in \( Q^A_\zeta \). By Cauchy, or sequential convergence, we mean a sequence of points \( \{ L_i \} \subseteq Q^A_\zeta \), such that, for all \( \varepsilon > 0 \), there exists \( k_\varepsilon \in \mathbb{N} \) such that \( k > k_\varepsilon \) implies \( d_\zeta (L_k, L_{k+1}) < \varepsilon \). Our strategy is to produce a Cauchy-convergent sequence of points which contains the non-convergent sequence \( L \) from the proof of Lemma 3.7 as a subsequence.

First, however, consider any two languages \( L \) and \( M \) in \( \mathcal{P}_A \). We will show that there is a language equidistant from \( L \) and \( M \). For each \( j \in \mathbb{N} \), select language \( N_{1,j} \subseteq (L \setminus M) \cap A^j \) such that \( \#N_{1,j} = \lfloor \frac{1}{2} \#((L \setminus M) \cap A^j) \rfloor \) and language \( N_{2,j} \subseteq (M \setminus L) \cap A^j \) such that \( \#N_{2,j} = \lfloor \frac{1}{2} \#((M \setminus L) \cap A^j) \rfloor \).

Define language \( N \) such that \( N[j] = (L \cap M) \cup N_{1,j} \cup N_{2,j} \) for all \( j \in \mathbb{N} \). If \( L \triangle M \) is an infinite language, it follows that \( d_\zeta (L, N) = d_\zeta (N, M) = \frac{1}{2} d_\zeta (L, M) \). If \( L \triangle M \) is finite, is true, as well, since all three distance are \( 0 \).

For \( i > 0 \), we obtain, by \( 2^k - 1 \) applications of the above conclusion, a sequence of languages \( N_{a,i} = \{ N_j \}_{j \leq 2^k} = \{ N_0 = J_{a,i}, N_1, \ldots, N_{2^{k-1}}, N_{2^i} = J_{a,i+1} \} \). The distance between language \( N_j \) and language \( N_{j+1} \) in \( N_{a,i} \), by Lemma 3.7, is

\[
\alpha^{-1} : 2^{-i+k}.
\] (3.14)

The concatenation of sequences \( N_{a,0}, N_{a,1}, \) and so on, gives a language sequence \( N'_a \) which contains \( N_a \) as a subsequence. Consider the point sequence \( L' = \eta_\zeta (N'_a) \). From (3.14) \( L' \) is Cauchy convergent but from Lemma 3.7 it is not convergent.

**Corollary 3.7.2** A language space is not compact under the Besicovitch topology.

**Proof.** Let \( \mathcal{O} \) be an open cover of \( \mathcal{P}_A \) defined by

\[
\mathcal{O} = \left\{ \{ M : d_\zeta (L, M) < \alpha^{-1} \} : L \in \mathcal{P}_A \right\}.
\]
Since $\alpha^{-1} < 2\alpha^{-1}$, we have, by Lemma 3.7, that any finite subset of $\mathcal{O}$ contains at most finitely many languages in $J_a$. Therefore, the open cover $\mathcal{O}$ has no finite subcover \cite{37}.

Of course, in settling this issue, we may no longer have the first condition for examining randomness on this space, namely, separability. Therefore, we seek a relation to a known compact space whereby we can guarantee separability.

3.3.5 An upper quotient space, homeomorphic to the unit interval

To obtain a compact space, both for the examination of randomness and to explore the most general features of the Besicovitch topology on language spaces, define the language norm $\|\cdot\|_{\zeta}$ as a quotient map from $Q^A_{\zeta}$ into the unit interval. This will result in three spaces: the non-$\text{T}_1$ language space under the topology induced by Besicovitch distance, the quotient space topologized by the metric quotient topology, and a compact upper quotient space with a well-known topology. We proceed formally as before, defining an equivalence relation $\equiv_{\zeta}$, the equivalence classes, and the quotient map which takes point in $Q^A_{\zeta}$ to their equivalence classes. We call the collection of equivalence classes the \textit{upper Besicovitch quotient space}, and we denote this space as $\mathcal{N}_{\zeta}$. In Theorem 3.1, we show that the space $\mathcal{N}_{\zeta}$ is homeomorphic to the unit interval (under the quotient topology on $\mathcal{N}_{\zeta}$).

Based on the formulation of the quotient space itself, we can summarize briefly the definitions involved.

**Definition 3.3.7** Suppose $L$ and $M$ are points in the Besicovitch quotient space, $Q^A_{\zeta}$. Then:

1. let $L \equiv_{\zeta} M$ if $\|L\|_{\zeta} = \|M\|_{\zeta}$ for all $L \in L, M \in M$;

2. let $(L)_{\zeta} = \{M \in Q^A_{\zeta} : M \equiv_{\zeta} L\}$, and denote by $\mathcal{N}_{\zeta}$ the collection $\{(L)_{\zeta} : L \in Q^A_{\zeta}\}$, denote elements of $\mathcal{N}_{\zeta}$ in script, $\mathcal{L}, \mathcal{M}, \mathcal{N}, \ldots$, and denote collections of elements of $\mathcal{N}_{\zeta}$ in boldface script, $\mathcal{L}, \mathcal{M}, \mathcal{N}, \ldots$; and

3. let $\kappa$ be the map from $Q^A_{\zeta}$ to $\mathcal{N}_{\zeta}$ which takes point $L$ to its equivalence class, $(L)_{\zeta}$.

Finally, for $r \in [0, 1]$, let $r_{\zeta}$ denote $\{L \in Q^A_{\zeta} : \|L\|_{\zeta} = r$ for all $L \in L\}$.

**Remark 7** It is obvious that $\equiv_{\zeta}$ is an equivalence relation. The quotient map $\kappa$ is well-defined by Corollary 3.5.3. Since $r_{\zeta} = (M)_{\zeta}$ for each $M \in r_{\zeta}$, this implies by Remark 6, that $r_{\zeta} = M$, for precisely one element $M \in \mathcal{N}_{\zeta}$. We equip the upper quotient space with a metric.
The upper quotient-space $Q_{\zeta}^A$ is the unit-interval, the Besicovitch quotient space is sphere-like, with antipodes and an “equator”, the set of points $L$, where $\kappa(L) = \frac{1}{2}$. The only compact equivalence classes, however, are the singletons, $0_{\zeta}$ and $1_{\zeta}$.

**Definition 3.3.8** Let the distance function $\rho : Q_{\zeta}^A \times Q_{\zeta}^A \to [0, 1]$ such that, if $\mathcal{L} = r_{\zeta}$ and $\mathcal{M} = s_{\zeta}$, for some $r, s \in [0, 1]$, then $\rho(\mathcal{L}, \mathcal{M}) = |r - s|$ as a metric on $Q_{\zeta}^A$. The collection of basis sets under the induced metric topology is the set

$$U = \{ \{ \mathcal{L} \subset Q_{\zeta}^A : r_{\zeta} \in \mathcal{L} \text{ if } |r - s| < \varepsilon \} : s \in [0, 1], \varepsilon > 0 \}.$$

**Remark 8** Then the set $U$ is apparently equivalent to the subset topology on the unit interval. That is, there is a homeomorphism between $Q_{\zeta}^A$ and $[0, 1]$ if the function $\rho$ induces the quotient topology on $Q_{\zeta}^A$.

We extend the abuse of notation used with languages and the quotient space and write $L \in r_{\zeta}$ (or $L \in \mathcal{L}$) to mean that language $L$ is to be found in points of the $\equiv_{\zeta}$ equivalence class $r_{\zeta}$ (or $L \in r_{\zeta} = \mathcal{L}$). We write $\mathcal{L} \subseteq r_{\zeta}$ to mean that each language in the language collection $\mathcal{L}$ is in one (but not necessarily the same) point in the $\equiv_{\zeta}$ equivalence class $r_{\zeta}$. We write $\mathcal{L} \subseteq \mathcal{L}$ to indicate that the image $\kappa[\eta_{\zeta}(\mathcal{L})]$ is a subset of the collection of elements of $\mathcal{L} \subseteq Q_{\zeta}^A$. We will show that, with exactly two exceptions, $r_{\zeta}$ is an uncountable subset of $Q_{\zeta}^A$. The elements $0_{\zeta}$ and $1_{\zeta}$ are the exceptions.
Lemma 3.8 The $\equiv_\zeta$ equivalence classes $0_\zeta$ and $1_\zeta$ are singletons in $\Omega^A_\zeta$.

Proof. The $\equiv_\zeta$ equivalence class $0_\zeta$ contains only the $\sim_\zeta$ equivalence class $\eta_\zeta(\Lambda)$, since $\|L\|_\zeta = d_\zeta(L, \Lambda)$. Thus, $L \in 0_\zeta$ implies $d_\zeta(L, \Lambda) = 0$, which implies $L \sim_\zeta \Lambda$.

On the other hand, suppose languages $L$ and $M$ and points $L$ and $M$ are such that $L \in 1_\zeta$ and $M \in 1_\zeta$. By Remark 6, $\|L^c\|_\zeta = \|M^c\|_\zeta = 0$, which we have just seen means $L^c \sim_\zeta M^c$. But since $L \setminus M = M^c \setminus L^c$, it is elementary that $L \Delta M = L^c \Delta M^c$. Therefore

$$d_\zeta(L, M) = \|L \Delta M\|_\zeta = \|L^c \Delta M^c\|_\zeta = d_\zeta(L^c, M^c) = 0.$$  

Hence, $L \sim_\zeta M$. As $L$ and $M$ were chosen arbitrarily, it follows that $L = M$ and that $1_\zeta$ contains a single point: the $\sim_\zeta$ equivalence class $\eta_\zeta(A^*)$. □

Since $1_\zeta$ is a singleton, given a point $L$, there is exactly one point in $\Omega^A_\zeta$ at distance 1 from $L$. We give a special name to a pair of points related in this way.

Definition 3.3.9 If $L, M \in \Omega^A_\zeta$ and $d_\zeta(L, M) = 1$, then points $L$ and $M$ will be called antipodes. Then we will write $L = M^c$.

Lemma 3.9 Every point $L \in \Omega^A_\zeta$ has a unique antipode in the (lower) Besicovitch quotient space.

Proof. From Corollary 3.5.1 this is equivalent to the claim that, if two languages are at distance 1 from the same language $L$ in point $L$, then they are $\sim_\zeta$-equivalent. But this is a consequence of the set theory identity

$$(L \Delta M_1) \Delta (L \Delta M_2) = M_1 \Delta M_2$$  

(3.16)

This provides the required condition, because if $d_\zeta(L, M_1) = 1$ and $d_\zeta(L, M_2) = 1$, this means that $L \Delta M_1$ and $L \Delta M_2$ are in $1_\zeta$ (from Definition 3.1.1), implying by Lemma 3.8 that

$$d_\zeta(L \Delta M_1, L \Delta M_2) = 0$$

and thus that $d_\zeta(M_1, M_2) = 0$, so $M_1 \sim_\zeta M_2$. □

Corollary 3.9.1 $L \in \mathcal{P}_A$, then $\|L^c\|_\zeta = 1 - \|L\|_\zeta$.
Corollary 3.9.2 \( L \in 0_\zeta \) if and only if \( L^c \in 1_\zeta \).

Corollary 3.9.3 \( \langle L^c \rangle_\zeta = \langle L \rangle_\zeta \), if and only if \( \|L\|_\zeta = \frac{1}{2} \) for any language \( L \in L \).

Definition 3.3.10 For each point \( L \in Q^A_\zeta \), the L-rotation of the point \( M \in Q^A_\zeta \) is the point \( \eta_\zeta (L \triangle M) \), for some language \( L \in L \). The L-rotation of \( M \) will be denoted \( \eta_{\zeta,L}(M) \). The L-rotation of the Besicovitch quotient space, denoted \( Q_{\zeta,L}^A \), is the collection \( \{ \eta_{\zeta,L}(M) : M \in Q^A_\zeta \} \).

The L-rotation of the \( \equiv_\zeta \)-equivalence class \( r_\zeta \), denoted \( r_{\zeta,L} \), is the set \( \{ M \in Q^A_\zeta : d_\zeta (M, L) = r \} \).

The L-rotation of the upper Besicovitch quotient space, i.e., the collection \( \{ r_{\zeta,L} : r \in [0, 1] \} \), will be denoted \( N_{\zeta,L} \).

Lemma 3.10 The L-rotation of the Besicovitch quotient space is equivalent as a set to the quotient space itself, and L-rotation is a bijection of the quotient space onto itself, and the resulting L-rotation of the upper quotient space is a bijection with the upper Besicovitch quotient space.

Proof. Well-definedness of \( \eta_{\zeta,L} \) as a function follows from the well-definedness of \( \eta_\zeta \). By Corollary 3.5.1 and the identity (3.16), \( \eta_{\zeta,L} \circ \eta_\zeta \) is an isometry between \( Q_A \) and \( Q_{\zeta,L}^A \). Therefore, if \( \eta_{\zeta,L}(M) = \eta_{\zeta,L}(N) \), then, if \( M \in M \) and \( N \in N \), \( M \sim_\zeta N \), so that \( \eta_{\zeta,L} \) is also injective. It follows that \( M \in r_\zeta \) if and only if \( \eta_{\zeta,L}(M) \in r_{\zeta,L} \). □

There are uncountably many \( \equiv_\zeta \) equivalence classes because the norm \( \| \cdot \|_\zeta \) is surjective on the unit interval. In addition, we now show that no open set in \( Q^A_\zeta \) is contained in a single \( \equiv_\zeta \) equivalence class. This is the essential condition for the proof that \( \rho \) is the quotient of \( d_\zeta \). We need the following straightforward proposition.

Proposition 2 If \( L \in Q_A \), \( \|L\|_\zeta = r \), and \( 0 \leq s \leq r (\leq 1) \), then there exists a subset of \( L \), the language \( M \subseteq L \), such that \( \|M\|_\zeta = s \).

Proof. If \( s = 0 \), let \( M = \Lambda \), and we are done. If \( r = s \), let \( M = L \), and we are done. Therefore, assume that \( s \in (0, r) \) to the cardinality of the set of shorter words.

Note that \( s/r > 0 \). Form the language sequence \( L = \{ L[i] \}_{i \in \mathbb{N}} \), and, from this, form the integer sequence \( \{ m_i \}_{i \in \mathbb{N}} \) such that the following equality holds:

\[
m_i = \left\lfloor (s/r) \#L[i] \right\rfloor.
\] (3.17)
Observe that for each $i \in \mathbb{N}$, $m_i \leq \#L[i] \leq \#A^i$. Then there exists a language sequence $\{M_i\}_{i \in \mathbb{N}}$ such that $M_i \subseteq L \cap A^i$ and $\#M_i = m_i$. Finally, let $M = \bigcup_{i \in \mathbb{N}} M_i$. By calculation,

$$0 \leq (s/r) \#L[<k] - \#M[<k] < k,$$

(3.18)

so

$$\|M\|_\zeta = (s/r) \|L\|_\zeta = s,$$

(3.19)

and $M \subseteq L$. This is what we needed. \hfill \square

**Remark 9** It follows that, if $0 \leq r \leq s \leq 1$, then for any language $L \in r_\zeta$, there exists language $M \supseteq L$ such that $M \in s_\zeta$. The appropriate language is $L$ if $0 = r = s$, $A^*$, if $s = 1$, and may be constructed as in 2 by inverting the fractions in 3.17, 3.18, and 3.19, if $s \in (0, 1)$.

**Lemma 3.11** No open set in the Besicovitch quotient space $Q^A_\zeta$ is a subset of a $\equiv_\zeta$ equivalence class.

**Proof.** Since $Q^A_\zeta$ is perfect, this is trivially true for the classes $0_\zeta$ and $1_\zeta$, by Lemma 3.8 and Corollary 3.6.1. Therefore, let $r \in (0, 1)$ and suppose language $L \in P_A$ and $L \in Q^A_\zeta$ such that $L \in L \in r_\zeta$. For any open set $L$ in $Q^A_\zeta$ containing $L$, there is a number $\varepsilon' > 0$ such that $d_\zeta (L, M) < \varepsilon'$ implies $M \in L$. We claim there exists a point $M \in L$ such that, for language $M \in M$, the norm of $M$ is distinct: $\|M\|_\zeta \neq \|L\|_\zeta$.

It is sufficient to exhibit a language $M \in M$ such that $\|M\|_\zeta \neq \|L\|_\zeta$ and $d_\zeta (L, M) < \varepsilon'$. First, let $\varepsilon = \min \{r/2, \varepsilon'/2\}$. Note that $\varepsilon' > \varepsilon > 0$. Our selection of $\varepsilon$ provides the following:

$$0 < \varepsilon < r \leq 1,$$

which implies that

$$0 < r - \varepsilon < r$$

(3.20)

Then, by Proposition 2, there is a language $M \subseteq L$ such that $\|M\|_\zeta = r - \varepsilon$. But since $r - \varepsilon < r$, $\|M\|_\zeta \neq \|L\|_\zeta$. Since $M \subseteq L$, it follows that

$$d_\zeta (L, M) = \|L \Delta M\|_\zeta = \|L \setminus M\|_\zeta.$$
However,
\[ \|L\|_\zeta = \|M\|_\zeta + \|L \setminus M\|_\zeta \]

from Corollary 3.1.1. Thus,
\[ d_\zeta(L, M) = \|L\|_\zeta - \|M\|_\zeta = r - (r - \varepsilon) = \varepsilon \] (3.21)

We have shown that there is a language \( M \) such that, if \( M \in L \) and \( \|M\|_\zeta \neq \|L\|_\zeta \). Since \( L \) was chosen arbitrarily, the conclusion is that no open set in the quotient space is contained in a \( \equiv_\zeta \) equivalence class. \( \square \)

**Corollary 3.11.1** If \( L \in \tau_\zeta \) is an open set in the Besicovitch topological language space, and language \( L \in L \), then there exists \( \varepsilon > 0 \) such that, for every real number \( r \in (\|L\|_\zeta - \varepsilon, \|L\|_\zeta + \varepsilon) \cap [0, 1] \),

there exists a language \( M \in L \) such that \( M \in r_\zeta \).

**Proof.** Assume \( L \in L \in \tau_\zeta \), as hypothesized. Then there exists \( \varepsilon > 0 \) such that \( d_\zeta(L, M) < \varepsilon \) implies \( M \in L \), by definition. In Lemma 3.11, it is in effect shown that if \( \|L\|_\zeta \in (0, 1) \), then, for all \( \varepsilon' > 0 \) such that \( \varepsilon' < \varepsilon \), there exists \( M \in L \) such that \( d_\zeta(L, M) = \varepsilon' \). But also selecting any \( \varepsilon'' \) such that \( \varepsilon'' < \min \left(1 - r, \frac{1}{2}, \varepsilon'/2\right) \), then \( r < r + \varepsilon'' < 1 \) and, by Remark 9, we conclude that, if \( \|L\|_\zeta \in (0, 1) \), the norms of languages in \( L \) include every value in some open subinterval of \([0, 1]\) (at the very least, the interval \((r - \delta, r + \delta)\), where \( \delta = \min \{\varepsilon', \varepsilon''\} \)) around \( \|L\|_\zeta \). The surjectivity of the norm allows us to find a similar half-open interval around any language \( L \in 0_\zeta \), and, by Lemma 3.9, the same is true for any language \( L \in 1_\zeta \). \( \square \)

This corollary asserts that, under the Besicovitch topology, representatives of some continuous interval of norm-values — the elements, that is, of some open set in the upper quotient space — can be found in every open set in the language space. In other words, as claimed in Remark 6(1), *the language norm \( \|\cdot\|_\zeta \) is a continuous map from \( \mathcal{P}_X \) onto \([0, 1]\).* However, Corollary 3.11.1 does not assert that all of the languages which map under the quotient mappings to an open set — in the
upper quotient space — are contained in every basis element of the pseudo-metric topology in the
language space. Indeed, we know that, in general, this is false. For instance, for every language \( L \) such that \( \| L \| \zeta = \frac{1}{2} \), every basis element of \( \tau \zeta \) except for \( \mathcal{P}A \) itself excludes every language \( M \) such that \( M \sim \zeta L^c \), i.e., uncountably many languages \( M \) such that \( \| M \| \zeta = \| L \| \zeta \).

**Theorem 3.1** The upper Besicovitch quotient space \( \mathcal{N} \zeta \) is homeomorphic to the unit interval \([0, 1]\).

In other words, the upper quotient space is the unit interval.

**Proof.** We claim that there is a bijection \( f \) between \( \mathcal{N} \zeta \) and \([0, 1]\) which is continuous and of which the inverse mapping \( f^{-1} \) is continuous, where \( \mathcal{N} \zeta \) is equipped with the quotient topology.

Consider the mapping \( f : \mathcal{N} \zeta \to [0, 1] \) such that \( f(L) = r \) if \( L = r \zeta \), which is shown to be well-defined in Remarks 7 and 8. Since all languages belonging to the same \( \sim \zeta \) equivalence class belong to a single \( \equiv \zeta \) equivalence class, \( f \) is injective. Also, the mapping \( f \) is surjective, since the norm \( \| \cdot \| \zeta \) is surjective by Corollary 3.2.1. Then \( f \) has a well-defined inverse, \( f^{-1} : [0, 1] \to \mathcal{N} \zeta \). We need to show that, if \( I \) is an open subset of \([0, 1]\), \( f^{-1}(I) = L \subset \mathcal{N} \zeta \) is open under the quotient topology on \( \mathcal{N} \zeta \), and, conversely, that, if \( L \) is open in \( \mathcal{N} \zeta \) under the quotient topology, then \( f(L) \) is open in \([0, 1]\).

Let \( \mathcal{L} \) be open. A set in \( \mathcal{N} \zeta \) is open under the quotient topology only if there is some open set \( \mathcal{L} \) of \( Q^A \) such that \( \kappa(L) = \mathcal{L} \), and, therefore, there is some open set \( L \) of \( \mathcal{P}A \) such that \( \eta(L) = \mathcal{L} \). If there is some open set \( \mathcal{L} \) of \( \mathcal{P}A \) such that \( \eta(L) = \mathcal{L} \), then, by Corollary 3.11.1, this means that for each language \( L \in \mathcal{L} \), there exists \( \varepsilon > 0 \) such that

\[
\left| \| L \| \zeta - \| M \| \zeta \right| < \varepsilon
\]

implies \( \| M \| \zeta \in \mathcal{L} \). But this means that, if \( \| L \| \zeta = r \), and \( | r - s | < \varepsilon \), then \( s \zeta \in \mathcal{L} \). For the converse, suppose that, for element \( r \zeta \) of \( \mathcal{L} \) there exists \( \varepsilon > 0 \) such that, if \( | r - s | < \varepsilon \) implies \( s \zeta \in \mathcal{L} \). Then the element \( s \zeta \) contains all languages with norm \( s \), i.e., it contains all the languages in \( \mathcal{P}A \) of norm \( s \). The basis element \( \mathcal{N} = \{ N : d \zeta (L, N) < \varepsilon \} \) of the Besicovitch topology is a subset of this collection. On the one hand, if \( N \in \mathcal{P}A \) such that \( d \zeta (L, N) < \varepsilon \),

\[
\varepsilon + \| N \| \zeta > d \zeta (L, N) + d \zeta (N, \Lambda) \geq d \zeta (L, \Lambda) = \| L \| \zeta = r,
\]

75
which means that
\[ \|N\|_\zeta > r - \varepsilon. \]

On the other hand,
\[ \|N\|_\zeta = d_\zeta (N, \Lambda) \leq d_\zeta (L, N) + d_\zeta (L, \Lambda) = d_\zeta (L, N) + \|L\|_\zeta < \varepsilon + r. \]

Hence, there is no language \( M \) in \( N \) such that
\[ \left| \|L\|_\zeta - \|M\|_\zeta \right| \geq \varepsilon. \] It follows that every set in \( \mathcal{N}_\zeta \) which is a union of finite intersections of sets of the form shown specified in equation (3.15) is open under the quotient of the Besicovitch topology.

The open sets in \([0, 1]\) are images under \( f \) of sets which are quotients in \( \mathcal{N}_\zeta \) of open sets in \( Q^A_\zeta \), and the quotients of open sets in \( Q^A_\zeta \) are sets \( \mathcal{N}_\zeta \) which have open images under \( f \). Thus, \( f \) is a homeomorphism with \( \mathcal{N}_\zeta \) as the quotient space of the Besicovitch topology on a language space. \( \square \)

This homeomorphism identifying the upper quotient space with the unit interval will be the key to understanding to randomness in the Besicovitch topology. The quotient map \( \kappa \) is of course not injective. For example, we have seen that all but perhaps finitely many sections of a word ideal, \( \{J_{w,i}\}_{i \in \mathbb{N}} \), belong to distinct \( \sim_\zeta \) equivalence classes, from equation (3.12), whereas, from equation (3.13), they all belong to the same \( \equiv_\zeta \) equivalence class.

**FACT 8** There exist right ideals of \( A^* \) in every \( \equiv_\zeta \) equivalence class of \( Q^A_\zeta \), by Lemma 3.5.

This is an observation regarding the location of languages in the upper quotient space, however. We consider the geometry of the situation more closely. A sketch of the relationship of the spaces and typical members of each is given in Figure 3.

### 3.4 The geometry of the Besicovitch topology

A language space is also lent a characteristic structure by the Besicovitch topology. It was shown that (only) the languages in \( L^c \) are at distance 1 from languages in the point \( L \) of \( Q^A_\zeta \). In subsection one, we reopen the discussion of languages in points of \( Q^A_\zeta \), to arrive at a more complete description of points. In subsection two, we develop a partial description of open and closed sets in the quotient space, and show how to produce an arbitrary boundary element of the closure of a basis sets of
Figure 3: The relationship of the Besicovitch language space, quotient space, and unit interval (upper quotient space).

Spaces and quotient maps are named (left); in the main part of the figure, the spaces are portrayed in oblong boxes (solid lines), the $\sim$ equivalence classes are enclosed in dotted lines, and the $\equiv$ equivalence classes are enclosed in dashed lines. The spaces $\mathcal{P}_A$, $\Omega_\zeta^A$, and $\mathcal{N}_\zeta$ are portrayed as extending from the element $0_\zeta$, containing only $\eta_\zeta(A)$, on the left, to the element $1_\zeta$, containing only $\eta_\zeta(A^*)$, on the right. One point $L \in \Omega_\zeta^A$ is portrayed shown to include the language $L$ (together with uncountably many others). One element $r_\zeta \in \mathcal{N}_\zeta$ is shown to include (uncountably many) points in $\Omega_\zeta^A$, among them the points $R$, $S$, and $T$, which are equivalence classes of languages.
the topology. In subsection three, we show an outline of the structure of an element in the upper
quotient space \( N_\zeta \), and show that, in general, the \( \equiv_\zeta \) equivalence classes are uncountable.

3.4.1 A point in the quotient space

We examine the relationship between languages \( L \) and \( M \) such that \( \eta_\zeta (L) = \eta_\zeta (M) = L \), where \( L \) is a point in \( Q^A_\zeta \). First, it is clear that \( L \sim_\zeta M \) if \( L \triangle M = \emptyset \) is a finite language. Equally clearly, this is not necessary for \( \sim_\zeta \) equivalence. The necessary condition is the converse of Lemma 3.4, which we state here for convenience.

**Lemma 3.12** \( L \sim_\zeta M \) if and only if, for all \( m \in \mathbb{N} \), there exists \( k'_m \in \mathbb{N} \) such that \( k > k'_m \) implies \( \# (L \triangle <k M) < \alpha^{-k} \), since \( \alpha^{-k} < \alpha^{-k-1} < \alpha^k \) for all \( k \in \mathbb{N} \).

We can exclude certain languages from \( 0_\zeta \). The above necessary condition is the reason that there are no ideals of \( \Lambda^* \) in \( 0_\zeta \). Ideals are a subset of a more general sort of language which cannot be found in \( 0_\zeta \). The ultimate conclusion is the following corollary.

**Corollary 3.12.1** If \( \alpha > 1 \) and \( L \in 0_\zeta \), then \( L^{[k]} = A^k \) for at most finite \( k \in \mathbb{N} \).

**Proof.** Toward contradiction, suppose there exists a language \( L \in 0_\zeta \) and a sequence of integers \( \{k_i\}_{i \in \mathbb{N}} \) such that \( L^{[k_i]} = A^{k_i} \) for all \( i \in \mathbb{N} \). But then, for each \( i \in \mathbb{N} \), \( \# L^{[<k_{i+1}]} > \alpha^{k_i} = \alpha^{(k_i+1)-1} \), so that, in terms of 3.12, there is no \( k'_i \), i.e., no integer such that the cardinality of words in \( \# L^{[<k]} \) is less than \( \alpha^{k-1} \), which is a contradiction. \( \square \)

This is a hint that there is a general difference between language spaces over different numbers of symbols under the Besicovitch topology, because, for any proper subset \( B \) of \( A \), then \( B^* \in 0_\zeta \), as will be shown in subsection 3.5.2.

**Corollary 3.12.2** If \( L \in 1_\zeta \), then there are at most finite \( k \in \mathbb{N} \) such that \( L^{[k]} = \Lambda \).

Recall the sequence \( K_\zeta (L, M) = \{k_m\}_{m \in \mathbb{N}} \) where, for each \( m \in \mathbb{N} \),

\[
k_m = \min \left\{ k'_m \geq m : k > k'_m \text{ implies } \# (L \triangle <k M) < \# A^{k-m} \right\},
\]

from Definition 3.3.3. Since \( k_m \geq m \) if \( K_\zeta (L, M) \) contains at least \( m + 1 \) terms, the sequence \( K_\zeta (L, M) \) either has finitely many terms or is unbounded.
Corollary 3.12.3  For \( L, M \in \mathcal{P}_A \), \( \mathcal{K}_\zeta (L, M) \) is a finite sequence if and only if \( L \not\sim \zeta M \). If \( L \sim \zeta M \), \( \mathcal{K}_\zeta (L, M) \) is strictly increasing.

Proof.  The first claim follows directly from Corollary 3.12. Assume, toward contradiction, that the second claim is false, i.e., that, for some \( m \in \mathbb{N} \), \( k_m = k_{m+1} \). Note that, in this case, \( k_m \neq m \), for this would imply \( k_{m+1} = m \), contrary to Definition 3.3.3. Therefore, \( \# (L \triangle ^{k_m-1} M) > \# A^{k_m-1} \), while \( \# (L \triangle ^{k_m+1} M) < \# A^{(k_m+1)-(m+1)} = \# A^{k_m-1} \). This is a contradiction, since \( (L \triangle ^{k_m+1} M) \supseteq (L \triangle ^{k_m} M) \). \( \square \)

In particular, if \( L \sim \zeta M \) and \( k_m = m \) for some \( m > 1 \), then \( k_i = i \) for all \( i \leq m \).

Lemma 3.13  The set difference \( L \triangle M \) is finite if and only if there exists \( n \in \mathbb{N} \) such that, for all \( m \in \mathbb{N} \), \( k_m \leq m + n \).

Proof.  (\( \Rightarrow \)) If \( L \triangle M \in \text{FIN}_A \) there exists \( n \in \mathbb{N} \) such that \( |w| \leq n \) if \( w \in L \triangle M \). Then \( \# (L \triangle ^{n+m} M) \leq \# A^{(m+n)-m} = \# A^n \) for all \( m \in \mathbb{N} \).

(\( \Leftarrow \)) Given \( n \) such that, for all \( m \in \mathbb{N} \), \( k_m \leq m + n \)

\[
L \triangle ^{k_m+1} M \leq \# A^{k_m+1-m} \leq \# A^{(m+n+1)-m} = \# A^{n+1}
\]

for all \( m \in \mathbb{N} \). Thus, \( L \triangle M \) is finite. \( \square \)

3.4.2  Open and closed neighborhoods of languages

The open neighborhoods in the pseudo-metric topology of the language space are generated by the basis elements \( B_\varepsilon(L) \), \( \varepsilon > 0 \), \( L \in \mathcal{P}_A \). If \( \varepsilon \geq 1 \), then \( B_\varepsilon(L) = \mathcal{P}_A \). If \( \varepsilon \geq \frac{1}{2} \), then \( B_\varepsilon(L) \) contains \( 0_\zeta \) or \( 1_\zeta \) depending upon whether \( \|L\|_\zeta \) is less than or greater than \( \frac{1}{2} \). Note the anomalous character of neighborhoods with radius greater than \( \frac{1}{2} \).

Example 11  Suppose language \( L \) is such that \( \|L\|_\zeta = \frac{5}{8} \). Then \( B_{11/16}(L) \) contains every language in both \( 0_\zeta \) and \( 1_\zeta \), but not every language \( M \) such that \( M = \frac{1}{2} \), since there is a point \( M \) containing a language \( N \) such that \( \|N\|_\zeta = \frac{1}{2} \) and \( d_\zeta(M, N) = \frac{1}{8} \) (by Proposition 2), but, considering the languages in \( M^c \), such as \( N^c \), the triangle inequality gives that \( d_\zeta(L, N^c) \geq 1-1/8 = 7/8 > 11/16 \). Hence, \( M^c \) is in the same \( \equiv_\zeta \) equivalence class as \( M \), which is only “1/8” away from the \( \equiv_\zeta \) class to which \( L \) belongs, but not in \( B_{11/16}(L) \).
Let $B_\varepsilon(L)$ denote the closed neighborhood of $L$ of radius $\varepsilon$. Closed neighborhoods contain all languages at distance $\varepsilon$ from language $L$, and these languages make up the boundary of the closed neighborhood. Since $\|L\|_\zeta = l \in [0, 1]$, the boundary of $B_\varepsilon(L)$ is then a subset of the set

$$\{ r_\zeta : \max \{ l - \varepsilon, 0 \} \leq r \leq \min \{ 1, l + \varepsilon \} \}.$$ 

A lower and upper end to the boundary may be discerned, the lower end being the set of languages in $0_\zeta$, if $l - \varepsilon < 0$, and among the languages with norm $l - \varepsilon$, otherwise. The upper end of the boundary, the set of boundary languages of greatest norm, lies within $1_\zeta$ if $l + \varepsilon \geq 1$, and among languages with norm $l + \varepsilon$, otherwise. We can also usually (i.e., under specific limitations on $\varepsilon$) locate elements of the boundary of a basis set with any norm between the upper and lower end of the boundary.

**Lemma 3.14** If $L \in \mathcal{P}_A$, $\|L\|_\zeta = l$, and $\varepsilon > 0$ such that the interval $[l - \varepsilon, l + \varepsilon]$ is a subset of the unit interval, and $\delta \in [l - \varepsilon, l + \varepsilon]$, then there exist two languages $S \subseteq L$ and $T \subseteq L^c$ such that

1. $\|S\|_\zeta = \frac{\varepsilon - \delta + l}{2}$, and
2. $\|T\|_\zeta = \frac{\varepsilon + \delta - l}{2}$.

and any language $M \sim_\zeta (L \setminus S) \cup T$ is such that $d_\zeta(L, M) = \varepsilon$ and $\|M\|_\zeta = \delta$.

**Proof.** Under the given conditions, $0 \leq l - \varepsilon < \delta < l + \varepsilon \leq 1$. This implies both that $\varepsilon + \delta > l$, so $\varepsilon + \delta - l > 0$, and that $\varepsilon - \delta + l > 0$. If $r = \frac{\varepsilon + l - \delta}{2}$ and $s = \frac{\varepsilon + l - \delta}{2}$, the $\equiv_\zeta$ equivalence classes $r_\zeta$ and $s_\zeta$ are non-empty. Furthermore, since $l + \varepsilon \leq 1$ and $\delta \leq 1$, $l + \varepsilon + \delta \leq 2$, so $\varepsilon + \delta - l \leq 2 - 2l$, and therefore

$$0 < \frac{\varepsilon + \delta - l}{2} \leq 1 - l = \|L^c\|_\zeta. \quad (3.22)$$

At the same time, since $\delta > \varepsilon$,

$$0 < \frac{\varepsilon - \delta + l}{2} < l = \|L\|_\zeta. \quad (3.23)$$

Applying Proposition 2 to both (3.22) and (3.23), we conclude that languages $S \subseteq L$ and $T \subseteq L^c$ can be constructed with the required norms. But then

$$d_\zeta(L, M) = \| (L \setminus M) \cup T \|_\zeta = \|L \setminus (L \setminus S)\|_\zeta + \|T\|_\zeta = \frac{\varepsilon - \delta + l}{2} + \frac{\varepsilon + \delta - l}{2} = \varepsilon,$$
and
\[ \|M\|_\zeta = \|L\setminus S\|_\zeta + \|T\|_\zeta = \left( l - \frac{\varepsilon - \delta + l}{2} \right) + \frac{\varepsilon + \delta - l}{2} = \delta. \]

**FACT 9** If \( l \leq \varepsilon \), no language \( M \) such that \( \left| l - \|M\|_\zeta \right| > \varepsilon \) is in \( \overline{B}_\varepsilon (L) \), and if \( \|L\|_\zeta = \|M\|_\zeta = r \), then \( d_\zeta (L, M) \leq 2 \cdot \min \{ r, 1 - r \} \).

**Proof.** The first claim is a restatement of what was shown in proving Theorem 3.1. The second claim follows from the triangle inequality and Corollary 3.9.1.

### 3.4.3 Ideals and the elements of the upper quotient space

We have seen evidence that right-sided ideals of \( A^* \) can be found throughout the Besicovitch topological language space. We develop this conclusion to a basic comprehension of elements of the upper quotient space. First, we extend the notion of sections of a word ideal to the \( n \)-word case.

**DEFINITION 3.4.1** An \( n \)-word ideal in the monoid \( A^* \) is a language \( J_F \) such that

\[ J_F = A^* w_1 A^* w_2 \cdots A^* w_n A^* \]

for some finite language \( F = \{ w_1, w_2, \ldots, w_n \} \) over \( A^* \). Then \( f_F = \sum_{i=1}^{n} |w_i| \) is the length of \( J_F \).

If \( v = (v_1, \ldots, v_n) \) is a vector over \( \mathbb{N}^{1 \times n} \), then the \( v \)-section of \( J_F \) is denoted \( J_{F,v} \) and is the right monoid ideal defined by

\[ J_{F,v} = A^{v_1} w_1 A^{v_2} w_2 A^{v_3} \cdots A^{v_n} w_n A^*. \]

Then the calculations involved in proving Lemmas 3.3 and 3.7 can be extended by induction, or else by the following.

**LEMMA 3.15** For every vector \( v \) over \( \mathbb{N}^{1 \times n} \) and every language \( F \) such that \( \#F = n \), the norm of the \( v \)-section of the \( n \)-word ideal \( J_F \) of length \( f_F \) is \( \alpha^{-f_F} \). That is, \( \|J_{F,v}\|_\zeta = \alpha^{-f_F} \).
Proof. Let \( v_1 + v_2 + \cdots + v_n = S \). Then, when \( k \geq f_F + S \), \( \# J_{F, v}^{[k]} = \alpha^{k-f_F} \). Therefore,

\[
\lim_{k \to \infty} \frac{\sum_{i=0}^{k-1} \alpha^{i-f_L} - \# J_{F, v}^{[k]}}{\sum_{i=0}^{k-1} \alpha^i} = \lim_{k \to \infty} \frac{\sum_{i=0}^{f_F+S-1} \alpha^{i-f_L}}{\sum_{i=0}^{k-1} \alpha^i} = \lim_{k \to \infty} \frac{\alpha^{s-1}}{\alpha^{k-1}} = 0,
\]

which implies that \( \| J_{F, v} \|_\zeta = \alpha^{-f_F} \). \( \square \)

From this we gather the notion that points in the upper quotient space contain languages that closely resemble unions of sections of monoid ideals of \( A^* \). They do not have to be such ideals; rather, cardinality of sections of these languages (as word-length goes to infinity) must approximate the cardinality of the unions of monoid ideals. Finally we show that all \( \equiv_\zeta \) classes except \( 0_\zeta \) and \( 1_\zeta \) are uncountable.

**Lemma 3.16** For any real number \( r \in (0, 1) \), the element \( r_\zeta \) of \( \mathcal{N}_\zeta \) is uncountable.

**Proof.** From Lemma 3.3 there is an \( r \)-simple language, \( L \). There exist at least two \( r \)-simple languages, though, since, for each \( r \in (0, 1) \),

\[
0 \leq \# L^{[k]} < \left[ r + \frac{1-r}{2} \right] \alpha^k = \left( \frac{r + 1}{2} \right) \alpha^k < r \alpha^k.
\]

This means that, for \( k \in \mathbb{N} \), there exists a subset of \( A^k \setminus L = (L^c)^{[k]} \) consisting of either \( \lfloor r \alpha^k \rfloor \) or \( \lfloor \left( \frac{1-r}{2} \right) \alpha^k \rfloor \) words, whichever is less, and a subset of \( L^{[k]} \) consisting of the same number of words.

This means there exists an \( r \)-simple language at distance

\[
s = \min \{ 2r, 1-r \}
\]

from \( L \). We construct this language as follows: let \( t_k = \min \{ \lfloor r \alpha^k \rfloor, \lfloor \left( \frac{1-r}{2} \right) \alpha^k \rfloor \} \) for \( k > N \); let \( T_k \) be a language such that \#\( T_k = t_k \) and \( T_k \subseteq (L^c)^{[k]} \), which is possible since \#\( (L^c)^{[k]} \geq 2t_k \); and let language \( F_k \) be a subset of \( L \) such that \#\( F_k = t_k \), which is possible since \( t_k \leq \# L^{[k]} \). Let \( T = \bigcup_{i \in \mathbb{N}} T_i \), \( F = \bigcup_{i \in \mathbb{N}} F_i \), and let \( N = L \setminus F \). Then the language \( L' \) defined by \( L' = N \cup T \) is the language formed by exchanging \( t_k \) words in \( L \) for \( t_k \) words in \( L^c \). Thus, the number of words
in $L \triangle^k L'$ is $2t_k = s\alpha^k$ for each $k \in \mathbb{N}$. Hence, $d_\zeta (L, L') = s$ and, since Land $L'$ contain the same number of words of each length, they have the same norm. Since $L$ is $r$-simple, so is $L'$.

For all $t \in \mathbb{R}$ such that $0 \leq t \leq s$, since $s \leq r$, there exists language $F' \subseteq F \subseteq L$ such that $\| F' \|_\zeta = t/2$, and language $T' \subseteq T = L' \cap L^c$ such that $\| T' \|_\zeta = t/2$ (by Proposition 2). Then, using Lemma 3.14 with $\varepsilon = t$ and $\delta = l = r$, language $L_t = (L \setminus F') \cup T'$ is such that $d_\zeta (L_t, L) = t$ and $\| L_t \|_\zeta = r$. Moreover, $L_t = (L \setminus F) \cup (F \setminus F') \cup T' = N \cup (F \setminus F') \cup T'$, so $L_t \triangle L' = (T \setminus T') \cup (F \setminus F')$. Thus, $d_\zeta (L_t, L') = s - t$.

This completes the proof. $\square$

We now consider what happens to the language families of the Chomsky hierarchy in the Besicovitch language topology.

3.5 The Chomsky hierarchy revisited

3.5.1 The finite languages are not dense

The example used to show that $d_\zeta$ is a strict pseudo-metric provides the clue to the simple proof that finite languages are of little significance in the Besicovitch topology. In fact, $\text{FIN}_A$ lies entirely within $0_\zeta$.

**Lemma 3.17** The finite languages are all in $0_\zeta$.

**Proof.** If a language $L$ is finite, there exists $N \in \mathbb{N}$ such that $n > N$ implies $L^{[n]} = \varnothing$, and hence that $\# (L \triangle^n \Lambda) = 0$. Therefore, for any finite language, $L$,

$$\| L \|_\zeta \leq \limsup_{k \to \infty} \frac{\sum_{j=0}^{N} \alpha^j}{\sum_{j=0}^{k-1} \alpha^j} = \limsup_{k \to \infty} \frac{\alpha^{N+1} - 1}{\alpha^k - 1} = 0.$$  

$\square$

In other words, precisely the opposite situation prevails in the Besicovitch topology from that in the Cantor topology: the finite languages are not dense, do not influence the norm of a language, and do not contribute to the distance between two languages. The next question must be: what if the
A word of length $k$ in a generally testable language with a set of permitted factors $S \subseteq A^n$. For this word, $k = qn + r$, where $q, r \in \mathbb{N}$ and $r < n$. Also, it can be seen that this is one of $s^\alpha r^\beta$ words of length $k$ in this language.

A description of a language is entirely finitary; that is, what if there is a finite language, a proper subset of $A^n$ for some $n \in \mathbb{N}$ which contains every factor, up to length $n$, of every word in the language?

### 3.5.2 All locally testable languages have norm zero

From the definition of a locally testable language, if $L \in LOC$ there is a fixed window length, $n \in \mathbb{N}$, such that by inspecting a word $w$ through a window which allows a view of only $n$ consecutive symbols of $w$, and by running this window over the word from one end to the other, one can determine whether $w$ belongs to $L$. We define a larger class of languages, *generally testable languages*, with the property that every locally testable language is a subset of some generally testable language. A language will be called generally testable if it includes every word in $A^*$ which is a concatenation of words from some subset of $A^n$, followed by an arbitrary factor of length less than $n$.

**Definition 3.5.1** A language $L$ is generally testable if there exists a window length $n \in \mathbb{N}$ and a set of permitted factors $S \subseteq A^n$, and $L = S^* A^{<n}$.

This means that word $w$ is in $L$ if and only if $w \in A^{<n}$ or $w$ can be written $u_1 u_2 \cdots u_t v$, where $u_i \in S$ for all $i \in \mathbb{N}_t$, $v \in A^{<n}$. This hardly seems to limit the size of a generally testable language, but the Besicovitch norm decisively differs from the cardinality-based notion of size.

**Lemma 3.18** Every generally testable language in $\mathcal{P}_A$ is in $0_\zeta$ except for $A^*$, which is in $1_\zeta$. 

84
Proof. We will take an arbitrary generally testable language \( L \) and show that its norm is either 1 or 0. Let \( L \) be a generally testable language with window-length \( n \) and permitted factors \( S \subseteq A^n \). Setting \#\( S \) = \( s \), first suppose that \( s = \alpha n \). This implies that \( S = A^n \), \( L = A^n \), and therefore that \( L \in 1 \). Therefore, suppose that \( s < \alpha n \). Given \( w \in L \), there exist unique \( q \) and \( r \), \( q \geq 0 \) and \( 0 \leq r < n \) such that \( |w| = nq + r \). By definition \( w = u_1 u_2 \cdots u_q v \), where \( u_i \in S \) for \( 1 \leq i \leq q \) and \( v \in A^r \). Calculating the number of words in \( A^n \) that are of the same length as \( w \), this means

\[
\#_L[|w|] = s^q \alpha^r
\] (3.24)

Therefore, writing, for convenience, \( q_i = \lfloor i/n \rfloor \) and \( r_i = i - kq_i \) for \( i \in \mathbb{N} \), the Besicovitch norm of \( L \) can be calculated as follows. Assume \( k' \geq n \). Then

\[
\|L\|_\zeta = \limsup_{k' \to \infty} \frac{\sum_{i=0}^{k'-2} s_{q_i} \left( \sum_{j=0}^{n-1} \alpha^j \right) + s_{k'-1} \sum_{j=0}^{k'-1} \alpha^j}{\#A^{k'-1}}, \quad \text{if } r_{k'-1} \neq 0, \text{ and }
\]

\[
\frac{\sum_{i=0}^{k'-1} s_{q_i} \left( \sum_{j=0}^{n-1} \alpha^j \right)}{\#A^{k'-1}}, \quad \text{if } r_{k'-1} = 0.
\]

This cannot exceed the limit of the sequence of ratios of word-count to possible words at length \( k' \), where \( k' = kq + n \), for, at these lengths, the cardinality of words possible in the last such \( n - 1 \) word-lengths \( k(q-1)+1, k(q-1)+2, \ldots, k(q-1)+n-1 \) is the maximum \( (\alpha^{n-1}) \). Consequently, with a change in index variable from \( k' \) to \( k = k'/n \), we have

\[
\|L\|_\zeta \leq \lim_{k \to \infty} \frac{\sum_{i=0}^{k} s^i \left( \sum_{j=0}^{n-1} \alpha^j \right)}{\alpha^{kn} - 1} \frac{\alpha - 1}{\alpha - 1} = \lim_{k \to \infty} \frac{s (\alpha^n - 1)}{s - 1} \lim_{k \to \infty} \frac{s^k - 1}{s^k} \frac{\alpha^{kn} - 1}{\alpha^{kn} - 1} = \frac{s (\alpha^n - 1)}{s - 1} \lim_{k \to \infty} \frac{s^k}{\alpha^{kn} + \frac{s^k}{\alpha^{kn} - 1}} = \frac{s (\alpha^n - 1)}{s - 1} \lim_{k \to \infty} \left[ \frac{s^k}{\alpha^{kn}} + \frac{s^k}{\alpha^{2kn} - \alpha^{kn}} - \frac{1}{s (\alpha^{kn} - 1)} \right]
\] (3.25)
and, since $\alpha^{2kn} - \alpha^{kn} \geq \alpha^{kn}$,

$$\|L\|_\zeta \leq \frac{2s(\alpha^n - 1)}{s - 1} \lim_{k \to \infty} \frac{s^k}{\alpha^{kn}}.$$  \hspace{1cm} (3.26)

Now note that since $s < \alpha^n$, $s/\alpha^n < 1$, and, since $s^k/\alpha^{kn} = (s/\alpha^n)^k$, it follows that the right hand side of (3.26), converges to zero. Thus $\|L\|_\zeta = 0$. \hspace{1cm} \Box

**Corollary 3.18.1** Every locally testable language belongs to $0_\zeta$.

**Proof.** Suppose language $L \in \text{LOC}_A$, with window length $k \in \mathbb{N}$ and permitted factors $S \subseteq A^k$. Consider the generally testable language $L'$ with the same window length and the same permitted factors. By the properties of a language norm, $\|L\|_\zeta \leq \|L'\|_\zeta$, and $\|L'\|_\zeta = 0$ from Lemma 3.18. \hspace{1cm} \Box

For example, if $\|M\|_\zeta = \frac{1}{2}$, then the basis element $B_{\frac{1}{4}}^\zeta(M)$ of $\tau_\zeta$, i.e.,

$$B_{\frac{1}{4}}^\zeta(M) = \left\{ N \in \mathcal{P}_A : d_\zeta(M, N) < \frac{1}{4} \right\}$$

contains no locally testable language. Therefore, the family $\text{LOC}_A$ is not dense in $(\mathcal{P}_A, \tau_\zeta)$.

### 3.5.3 Regular languages

All finite and local languages belong to $0_\zeta$. These sub-families of the regular languages contain no open set, nor does their closure contain any open sets in any of the three topological spaces associated with Besicovitch distance. Regular languages are possibly dense in the language space, and they are certainly to be found in elements of the upper quotient space at an arbitrarily small distance from any element in the upper quotient space.

**Lemma 3.19** Regular languages are dense in the upper Besicovitch quotient space $(\mathcal{N}_\zeta)$.

**Proof.** Let $r \in [0, 1]$. The claim is that, for all $\varepsilon > 0$, there exists a regular language $L$ such that $\|L\|_\zeta - r < \varepsilon$. If $\varepsilon \geq \min \{r, 1 - r\}$, either $\Lambda$ or $A^*$ satisfies the claim and we are done. Assume therefore that $\varepsilon < \min \{r, 1 - r\}$. Then $r < r + \varepsilon < 1$. Let integers $n$ and $q$ be such that $r < q\alpha^{-n} \leq r + \varepsilon \leq (q + 1)\alpha^{-n}$, and $0 < q < \alpha^n$. From this we have

$$0 < q\alpha^{-n} - r < \varepsilon.$$  \hspace{1cm} (3.27)
Let the language $S_\varepsilon$ be a subset of $A^n$ of cardinality $q$. Consider the right monoid ideal $S_\varepsilon A^*$, which is a disjoint union of the $q$ right word ideals $wA^*$, where $w \in S_\varepsilon$. Note that each of these is a 1-word ideal (a right word ideal) section $J_{F,v}$, where $F = \{w\}$ for the word $w \in S_\varepsilon$ and $v = (0)$.

Therefore, by Lemmas 3.1.1 and 3.15,

$$\|S_\varepsilon A^*\|_\zeta = \sum_{w \in S_\varepsilon} \|wA^*\|_\zeta = q\alpha^{-n}$$

From (3.27), this means that $\|S_\varepsilon A^*\|_\zeta - r < \varepsilon$, as required. Finally, by the Myhill-Nerode Theorem, $S_\varepsilon A^*$ is a regular language, since all but finitely many words in $S_\varepsilon A^*$ can be followed by $A^*$. $\square$

This means that the linear, context free, context sensitive and recursively enumerable languages are all dense in the upper quotient space. It does not, however, inform us about where these families lie in the language space or in the space $Q^A_\zeta$. This leads to the following corollary and a conjecture as to the situation in the language space.

**Corollary 3.19.1** If $r = q\alpha^{-n}$, for some pair $q, n \in \mathbb{N}$, then there is a regular language with norm $r$.

**Proof.** Trivial, as shown in the proof of Lemma 3.19. $\square$

The following, unfortunately, must stand for the time being as a conjecture. The difficulty is that the portion of the language which we use to define the “regularity” in a regular language is necessarily finite, i.e., some particular section of the language. The Besicovitch distance, as we have seen disregards finite subsets of a language.

**Conjecture 1** The regular languages are dense in the Besicovitch topology.

The non-RE languages, however, are ubiquitous.

3.5.4 Non-RE languages

As we will readily show, there is a non-RE languages in every point of $Q^A_\zeta$. This is a consequence of the not-surprising fact that, because $d_\zeta$ is a strict pseudo-metric, the $\sim_\zeta$ equivalence classes are uncountable.

87
Lemma 3.20  The single element of the class $0_\zeta$ is uncountable in $\mathcal{P}_A$ and contains a non-RE language.

Proof. Suppose that symbol $a \in A$. In $\mathcal{P}_A$, if a language $L$ is in $0_\zeta$ it may be easily verified that, for all $A$, the sequence $\#L^{\leq k}/(k-1)$, $k \geq 1$ converges to zero. For example, if $L = \{a^{n_k} : k \in \mathbb{N}\}$, for some $n > 1$, then the sequence $\left\{\frac{\#L^{\leq k}}{k-1}\right\}_{k \in \mathbb{N}}$ is comparable to the convergent sequence $\left\{\frac{\log n}{k}\right\}_{k \in \mathbb{N}}$, so that $\|L\|_\zeta = 0$. Let the integer sequence $\{k_i\}_{i \in \mathbb{N}}$ be a strictly increasing, non-negative sequence that is not recursively enumerable. Then the language $L' = \{a^{2k_i} : i \in \mathbb{N}\}$ is non-RE, and $L' \subseteq L$. Moreover, there are uncountably many non-RE sequences $\{k_i\}_{i \in \mathbb{N}}$. □

Corollary 3.20.1  Every $\sim_\zeta$ equivalence class contains a non-RE language.

Proof. Let $L$ be a point in $Q^A_\zeta$, and let language $L \in L$. If $L$ is non-RE, there is nothing to prove. Otherwise, consider any non-RE language $M \in 0_\zeta$. The union $L \cup M$ is in Land $L \cup M$ is non-RE. For, suppose that $L \cup M$ is RE. Then $L \setminus (L \cup M)$ is non-RE, which is impossible. □

These results suggest that randomness in the Besicovitch topology may be less definitive than in the Cantor topology. That is, finite languages are confined to a single equivalence class, regular languages are in a of set of $\equiv_\zeta$ equivalence classes, and non-RE languages are literally everywhere. To test the possibility that the non-RE/RE split in the Cantor topology is weakened here, we apply the approach outlined in Subsection 1.5.2.

3.6 Random Languages under the Besicovitch pseudo-metric topology

The space $(\mathcal{P}_A, \tau_\zeta)$ is not compact and perhaps non-separable. The same is true of the quotient space $Q^A_\zeta$. However, the upper quotient space $\mathcal{M}_\zeta$ is homeomorphic to $[0, 1]$, and is compact and separable. Let the topology $\tau_u$ be the subset topology induced on the unit interval by the usual Euclidean metric topology on $\mathbb{R}$. Then the set $B_u = \{(n\alpha^{-m}, (n + 1)\alpha^{-m}) \cap [0, 1] : n \in \mathbb{Z}, m \in \mathbb{N}, m > 0\}$ forms a countable, and hence enumerable basis for $\tau_u$. It follows that $([0, 1], B_u, \mu_u)$ is a randomness space, where $\mu_u$ is the Lebesgue measure, a probability measure on the unit interval. This space, historically, is one of the first that were examined using randomness tests (Example 3.6.4, [16]). In our context, the most useful single result is the fact that computable real numbers in the unit interval are nonrandom.
In the randomness space \(([0, 1], B_u, \mu_u)\), a real number is nonrandom if it belongs to the intersection of a nested set of computable unions of elements of \(B_u\), where the total length of the sets in subsequent unions goes to zero rapidly. Lebesgue measure is upper semi-computable, by summing the measure of sets of measure \(\alpha^{-1}, \alpha^{-2}, \alpha^{-3}, \ldots\). This means the conclusions of Theorem 1.2 and 1.3 on page 33 hold on \(\mathcal{N}_\zeta\).

**Definition 3.6.1** A point \(r_\zeta \in \mathcal{N}_\zeta\), and every language \(L\) in \(r_\zeta\), will be called nonrandom if \(r\) is nonrandom in \(([0, 1], B_u, \mu_u)\), and \(r_\zeta\), together with every language \(L \in r_\zeta\), will be called random otherwise.

The question then remains: what is a random number, by this method? A computable number, by the following definition, is immediately identifiable as a nonrandom real number.

**Definition 3.6.2** A real number \(r\) is computable if there exists a Turing Machine which can generate the \(\beta\)-ary expansion of \(r\), for any \(\beta \in \mathbb{N}\) where \(\beta > 1\).

For example, every algebraic and certain non-algebraic numbers such as \(\pi\) are computable, and hence nonrandom, by the following lemma.

**Lemma 3.21** [Example 3.61, [16]] Every computable number is nonrandom.

**Proof.** The machine which generates the \(\alpha\)-ary expansion of a computable number \(q\) successively confines the value of \(q\) to an interval of \(\mu_u\)-measure \(\alpha^{-1}, \alpha^{-2}, \ldots\). This is a randomness test on \(([0, 1], B_u, \mu_u)\) containing the number \(q\). \(\square\)

If, for example, the expansion of the real number \(q\) to three places is \(0.d_1d_2d_3\), then

\[
q \in (d_1\alpha^{-1}, (d_1 + 1)\alpha^{-1}) \cap ((d_1\alpha + d_2)\alpha^{-2}, (d_1\alpha + d_2 + 1)\alpha^{-2}) \cap ((d_1\alpha^2 + d_2\alpha + d_3)\alpha^{-3}, (d_1\alpha^2 + d_2\alpha + d_3 + 1)\alpha^{-2}) \cap \cdots,
\]

which is clearly a randomness test.

From Definition 3.6.1, every language in \(0_\zeta\) is nonrandom in the Besicovitch topology. Combining this with Lemma 3.20, it emerges that there is a nonrandom language that is non-RE, which is to say, the set of random languages under Besicovitch and Cantor topologies do not coincide.
A fortiori there is a $\equiv_\zeta$ equivalence class containing nothing but non-RE languages, since the unit interval is uncountable and RE languages are countable. However, we can sharpen this observation from the standpoint of randomness.

**Lemma 3.22** There is a $\equiv_\zeta$ equivalence class $r_\zeta$ for some $r \in [0, 1]$ which contains no RE languages and in which every language in $r_\zeta$ is random.

**Proof.** We proceed by contradiction. Let $r$ be a random element of $([0, 1], \tau_u, \mu_u)$. Then $r$ is not computable, by Lemma 3.21. Suppose, toward contradiction, that $L$ is an RE language such that $\|L\|_\zeta = r$. Since $L$ is RE, the characteristic function of $L$ is computable, which is to say, there is a Turing Machine which, given $w \in P_A$, computes $\chi_L(w)$, where $\chi_L : A^* \to \{0, 1\}$ such that

$$\chi_L(w) = \begin{cases} 
0, & w \notin L \\
1, & w \in L.
\end{cases}$$

Therefore, the functions $\{f_i\}_{i \in \mathbb{N}}$ are all computable, where

$$f_i(L) = \sum_{w \in A^i} \chi_L(w).$$

Thus the recursive function $S : \mathbb{N} \to \mathbb{R}$ defined by $S(k) = \sum_{i=0}^{k} f_i(L) / \#A^\leq k$ computes the norm of $L$, giving the $\alpha$-ary expansion of $r$. Hence, meaning that $r$ is computable, in contradiction to our assumption.

The conclusion is that there is no RE language in $r_\zeta$. □

Another conclusion may be drawn from Theorem 1.3 on page 33. First, the set of nonrandom points in the unit interval is of $\mu_u$-measure 0. However, by Lemma 3.16, we also see that uncountably many non-RE languages are found among the nonrandom languages (a more precise justification of this will be put forward in Chapter 5, in the discussion following Lemma 5.3 on page 117). The rational numbers are nonrandom and dense in the unit interval, hence, by Lemma 3.11, every open set in the Besicovitch upper quotient space maps under $\kappa$ to a set containing a nonrandom number, and that every open set in $P_A$ maps under $\eta_\zeta \circ \kappa$ to a set containing a nonrandom number. Therefore, by Theorem 1.3, since every open set in $P_A$ contains a nonrandom language, the random languages are meager in $P_A$. 90
Theorem 3.22.1 The RE, CS, CF, LIN, and REG families are nowhere dense in a language space under the Besicovitch language topology.

Proof. By Lemma 3.19, REG and hence all the other families listed are dense in $\mathcal{N}_\zeta$. The sequence used in the proof of Lemma 3.22 establishes that these families are nonrandom. By Theorem 1.3, it follows that nonrandom languages, being dense in the space $\mathcal{N}_\zeta$, are nowhere dense in $\mathcal{N}_\zeta$. By the fact that this is a quotient space of $\Omega_A^\zeta$, which is, in turn, a quotient space of the language space, these families are nowhere dense in $\mathcal{P}_A$ under the Besicovitch language topology. For we have that every open set in $\mathcal{P}_A$ maps under $\kappa \circ \eta_\zeta$ to an open set in $\mathcal{N}_\zeta$, by Lemma 3.11 and Corollary 3.11.1. Since in $\mathcal{N}_\zeta$ there is no open set, the closure of which contains only nonrandom elements, there can be none in the language space itself.

Returning to the original notion of randomness, i.e., that random elements of a space are the typical elements, we conclude that, under the Besicovitch topology, not only is the “average” language non-RE, but, even in the most selective and exceptional of language sets, one usually encounters non-RE languages.

In a sense, this pseudo-metric has overshot the mark when it comes to a topological distinction between languages, in separating all locally testable languages into a single equivalence class. Yet the Chomsky hierarchy is dense in the quotient space. We present in the next chapter a pseudo-metric closely tied to traditional interests in both language theory and information theory. It is adapted from the notion of the topological entropy of languages.
Chapter 4
The entropic pseudo-metric

Our third pseudo-metric begins by adapting a primary concept from formal language theory, symbolic dynamics, and information theory. In doing so, we build into the topology the ability to discern certain characteristics of languages. Much of the work done in the previous chapter can be applied to this topology, but not to the Cantor topology, showing that the two pseudo-metrics belong to general sub-class, where the language norm is a bijection with a connected set. The notion of distance involved here, though, does not involve a cumulative evaluation of the symmetric set-difference of two languages but an estimate of the rate of exponential growth, with respect to word length, of that difference. If a language is factorial, this is the rate of exponential growth in the number of factors in the language comprising the symmetric set-difference of two languages, with respect to the length of the factor. But this is precisely the topological entropy of the language. The word “topological” relates this quantity to the shift operation, which is a topological conjugacy on infinite sequences under the Cantor topology. Our modification of the topological entropy function into a language pseudo-metric involves three alterations: first, we normalize the exponential growth rate with respect to the number of symbols in the alphabet to bound it above by one; secondly, we exclude the possibility of an entropy of \(-\infty\), instead bounding the function below by zero; third, and most significant, we do not (necessarily) count the factors of words in the language, but only concern ourselves with the exponential rate of growth of the number of words in the language with respect to the number of symbols in the words.

We call this entropic distance. Like Besicovitch distance, entropic distance is a strict pseudo-metric, since many distinct languages have the same entropy. The pseudo-metric topology turns out to be a refinement of the Besicovitch topology, and hence non-compact, as well. The languages of norm less than one turn out, in fact, to be a subset of the languages of Besicovitch norm 0.

Entropy originally finds a place in information theory as a measurement of the relative amount of information a channel can carry under a particular encoding, where this is evaluating in terms
of the number of symbols necessary encode a message. In dynamics, entropy is a measure of the randomness of a dynamical system[23]. Finally, on a subshift of finite type, manifested in a factorial, prolongable, regular language, topological entropy measures the exponential growth rate of the number of blocks of symbols. Without going into these approaches, we can say that the use of topological entropy in the measurement of the distance between languages agrees with the intuitive notion that, the farther two languages are from each other, there more decisive should be the increase in the number of words in one, but not the other as word-length increases. However, in the application of entropy to language distance, the propagation of factors in the symmetric set-difference of the two is not necessarily significant, because we have no way, a priori, of distinguishing accidental occurrences of symbols from accidental occurrences of meaningful factors, and in distinguishing accidental occurrences of meaningful factors from meaningful and integral occurrences of such factors. For example, in the English word “compact”, the factors “mpa”, “act”, and “com”, upon closer acquaintance with the language, fall into the three categories mentioned. We would rather say, therefore, that, in our approximation to distance between languages, we will give weight solely to those sequences of symbols which are unquestionably elements of the language, i.e., to entire words.

Where the Cantor metric $d_1$ hinges on the “first apparent distinction” between languages, and disregards the rest, and the Besicovitch distance $d_\zeta$ relies upon the “total proportion” of the monoid exhausted by language distinctions, regardless of the expansion rate of this proportion, the entropic distance $d_h$ will take into account both the appearance of distinctions and their rate of increase. We will show that there exist locally testable languages with non-zero entropic language norms. We will show that the entropic language topology is a strict refinement of the Besicovitch language topology. In section one, we define the entropic pseudo-metric and norm and the quotient spaces we will employ in their investigation. In section two, we discuss the Chomsky hierarchy under the entropic topology. In section three, we sketch out the limited notion of randomness possible using the means established in Chapter Three.

4.1 Definition of entropy and entropic distance

Suppose that, given a word in a language, there is reasonable certainty that a selection of symbols of a certain cardinality can be appended to this word resulting in another word in the language.
Suppose further that there is a bijection between the set of symbols so appended, and a second set of symbols we can append to the new words in the language such that we will arrive, with reasonable certainty, at a new set of words in the language. Continuing in this way, we can forecast the exponential growth of the number of words in the language with respect to the linear increase in word-length. If the language in question is the symmetric set-difference of two languages, then this growth expresses a sort of distance between the languages. In the first subsection, we define the entropic language pseudo-metric and language norm to quantify the type of expansion described. In the second subsection, we show that the entropic language norm is surjective on the unit interval.

4.1.1 Entropy: the rate of exponential language growth

If the exponential growth in the number of words in a factorial language \( L \) were constant with respect to word-length, then for each \( k \in \mathbb{N} \) the quantity

\[
\frac{1}{k} \log \#L[k]
\]

would give this rate. Topological entropy takes the limit, as \( k \) grows without bound, of the above function over all factors of words in the language. Thus, it is given by a function \( h : \mathcal{P}_A \to \mathbb{R} \) such that

\[
h(L) = \limsup_{k \to \infty} \left( \frac{1}{k} \right) \log \#F_L[k],
\]

where \( F_L = \{ w \in \text{Fac}(v) : v \in L \} \). Since we have assumed \( L \) is a factorial language, \( F_L = L \), and

\[
h(L) = \limsup_{k \to \infty} \left( \frac{1}{k} \right) \log \#L[k].
\]

The base of the logarithm in (4.1) is assumed to be 2. To normalize the function \( h \), we modify the base of the logarithm and also bound it below by 0, resulting in a function which maps \( \mathcal{P}_A \) into \([0, 1]\).

**Definition 4.1.1** For languages \( L \) and \( M \) over alphabet \( A \), where \( \#A = \alpha \), we will denote by \( \| L \|_h \) the entropic norm of language \( L \), and by \( d_h(L, M) \) the entropic distance between \( L \) and \( M \),
where functions $\| \cdot \|_h : \mathcal{P}_A \to \mathbb{R}^{\geq 0}$ and $d_h : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}^{\geq 0}$ are such that

$$
\| L \|_h = \limsup_{k \to \infty} \frac{1}{k} \max \left\{ 0, \log_\alpha \# L[k] \right\}, \quad \text{and}
$$

$$
d_h(L, M) = \| L \triangle M \|_h = \limsup_{k \to \infty} \frac{1}{k} \max \left\{ 0, \log_\alpha \# (L \triangle M[k]) \right\}.
$$

**Remark 10** Then the function $\| \cdot \|_h$ is indeed a language norm, since $\| \Lambda \|_\zeta = 0$ and, if $L \subseteq M$, then $\# L[k] \leq \# M[k]$ for all $k$; this, in turn, means that the function $d_h$ is a language pseudo-metric.

**Remark 11** If there exists a sequence of integers $\{k_i\}_{i \in \mathbb{N}}$ such that $\# (L \triangle k_i M) = \alpha^{k_i}$ for all $i \in \mathbb{N}$, then $\lim_{i \to \infty} \frac{1}{k_i} \max \left\{ \log_\alpha \# (L \triangle k_i M), 0 \right\} = 1$, and languages $L$ and $M$ are at entropic distance 1. Recall that the same condition in the Besicovitch language space merely means that $d_\zeta(L, M) > 0$.

**Remark 12** Except when the language $L$ is finite, the entropic norm is proportional to the topological entropy of an infinite factorial language. Indeed, $\| L \|_h = \frac{h(L)}{\log_2 \alpha}$, if $L$ is a factorial language. In general, however, the following inequality holds:

$$
\| L \|_h \leq \frac{h(L)}{\log_2 \alpha}.
$$

Note that the permutative version of the entropic distance, namely,

$$
d_h^{(\alpha)}(L, M) = \limsup_{k \to \infty} \frac{1}{k} \max \left\{ \log_\alpha \min_{\pi \in S_\alpha} \left\{ \# (L \triangle \pi[k] M) \right\}, 0 \right\},
$$

is also a strict pseudo-metric. It is obvious that $d_h^{(\alpha)}(L, M) \leq d_h(L, M)$. As with the permutative version of the Besicovitch distance $d_\zeta^{(\alpha)}$, to make this well-defined, we must use the $\pi$-difference of least norm. Recall that the bracketed superscript “[k]” denotes the $\pi$-difference is of words of length $k$. Because of this, the limit may be realized over varying permutations at various word-lengths. This lends the permutative version of entropic distance a complexity which requires further research.

Some elementary conclusions are true for $d_h$ just as they are for $d_\zeta$. This is, in part, due to features explicitly design into the two pseudo-metrics, in part a result of the common range that they share, and also some of the of general properties of any language norm and language pseudo-metric. For
an example of the latter, no two languages can be farther apart than the language $A^*$ and $\Lambda$. From Remark 11, these languages are at distance 1.

**Lemma 4.1** The entropic distance between a language and the complement of that language is 1.

**Proof.** Trivial, from Remark 11. □

On any language space $\mathcal{P}_A$, entropic distance is a strict pseudo-metric. For instance, $\|\Lambda\|_h = \|\{a\}\|_h = 0$, where symbol $a \in A$. We will settle, in the following subsection, the question of the image of an entire language space under the entropic language norm.

### 4.1.2 The entropic language norm is surjective.

It is to be suspected that the entropic norm would be surjective, since entropy is known to have irrational values. Proportionality of the norm as mentioned in Remark 12 would hardly affect this! A construction slightly different from that used to establish the surjectivity of the Besicovitch language norm is required to demonstrate surjectivity of the entropic language norm.

**Definition 4.1.2** Given $r \in [0, 1]$, a language $L \in \mathcal{P}_A$ will be called $r$-expansive if, for each $k \in \mathbb{N}$,

$$\#L'[k] = \left\lfloor (\alpha^r)^k \right\rfloor.$$

The collection of all $r$-expansive languages will be denoted $L'_r$.

**Lemma 4.2** The entropic language norm is surjective, i.e. $\|\cdot\|_h : \mathcal{P}_A \rightarrow [0, 1]$.

**Proof.** We claim, first, that an $r$-expansive language exists for each $r \in [0, 1]$, and, secondly, that the entropic norm of an $r$-expansive language is $r$. Let $r \in [0, 1]$. If $r = 1$ or $r = 0$, then $\|\Lambda\|_h = 0$ or $\|A^*\|_h = 1$, which, in each case, satisfies both claims at once. Therefore, assume $r \in (0, 1)$.

Toward the first claim, since $\alpha^r < \alpha$ if $r < 1$, we conclude $\left\lfloor (\alpha^r)^k \right\rfloor < \alpha^k$. Therefore, an $r$-expansive language can be constructed. The claim is established.

Toward the second claim, assume that $L' \in L'_r$. We have the following.

$$\limsup_{k \to \infty} \frac{1}{k} \log_\alpha \#L'[k] = \|L'\|_h \leq \limsup_{k \in \mathbb{N}} \frac{1}{k} \log_\alpha (\alpha^r)^k \leq \limsup_{k \to \infty} \frac{1}{k} \log_\alpha \left(\#L'[k] + 1\right) \quad (4.4)$$
where
\[
\limsup_{k \to \infty} \frac{1}{k} \log_\alpha (\alpha^r)^k = \lim_{k \to \infty} r = r,
\]
and
\[
\lim_{k \to \infty} \frac{1}{k} \left[ \log_\alpha \left( \#L'[k] + 1 \right) - \log_\alpha \#L'[k] \right] = \lim_{k \to \infty} \frac{1}{k} \log_\alpha \left( \frac{\#L'[k] + 1}{\#L'[k]} \right) = 0
\]
so that, for all \( \varepsilon > 0 \), there exists \( k_\varepsilon \) such that \( k > k_\varepsilon \) implies
\[
\lim_{k \to \infty} \frac{1}{k} \left[ \log_\alpha \left( \#L'[k] + 1 \right) - \log_\alpha \#L'[k] \right] < \varepsilon,
\]
i.e.,
\[
\limsup_{k \to \infty} \frac{1}{k} \log_\alpha \#L'[k] + \varepsilon \geq \lim_{k \to \infty} \frac{1}{k} \log_\alpha \#L'[k] + \varepsilon
\]
\[
\geq \lim_{k \to \infty} \frac{1}{k} \log_\alpha \left( \frac{\#L'[k] + 1}{\#L'[k]} \right) = r,
\]
or
\[
\|L'\|_h + \varepsilon \geq r,
\]
implying that \( r \not\approx \|L'\|_h \). Together with (4.4), this means that \( \|L'\|_h = r \). The proof is complete. □

This means that every conclusion in Chapter Two which depended solely upon the surjectivity of the Besicovitch norm can be translated into a conclusion regarding the entropic norm. It also gives evidence of a fundamental distinction between entropic and Besicovitch distance, as we show in the next section. Prefatory to that, we need a metric quotient space through which to comprehend the pseudo-metric topology on \( \mathcal{P}_A \).

4.2 The entropic quotient space

Like \( d_\zeta \), the pseudo-metric \( d_h \) induces a topology on \( \mathcal{P}_A \) with a metric quotient topology in a quotient space above \( \mathcal{P}_A \). We denote this entropic quotient space \( \Omega^A_h \). A second, upper quotient space, denoted \( \mathcal{N}_h \), is induced on \( \Omega_h \) by the norm \( \|\cdot\|_h \). The definitions of these spaces and the distances upon them, the metrics \( d_\zeta \) and \( \rho \), are given in subsection one, together with the proof that the quotient map from \( \mathcal{P}_A \) to \( \Omega^A_h \) is an isometry. The pseudo-metric topology on \( \mathcal{P}_A \) has a quotient metric topology on \( \Omega^A_h \). In the second subsection, there is an analysis of a point in
the quotient space, which is the same thing as the determination of when two languages are at the entropic distance of 0. All of these spaces, distances, and relationships parallel the similarly-denoted structures derived for the investigation of $d_\zeta$. The conclusion in the second subsection clarifies that the entropic topology is a refinement of the Besicovitch topology. The third subsection, by pointing out the relationship between the lower and upper entropic quotient space, establishes what sort of refinement the entropic topology is of the Besicovitch topology. Namely, all languages of entropic norm less than one are found in the point $0_\zeta$ of $\Omega_\zeta^A$. We show that there is no homeomorphism between $\mathcal{A}_h$ and the unit interval. Thus, the randomness discussed in the next section cannot be understood in the same way as that established for the space $(\mathcal{P}_A, \tau_\zeta)$.

4.2.1 Definition of the entropic quotient spaces

Note that by exchanging subscripts we can use the notation employed with Besicovitch distance, deriving an equivalence relation, quotient space, second equivalence relation, upper quotient space, and, ultimately, the induced entropic topology of language spaces.

**Lemma 4.3** The relation $\sim_h$ defined by $L \sim_h M$ if $d_h(L, M) = 0$ is an equivalence relation on $\mathcal{P}_A$.

**Proof.** Trivial. □

**Definition 4.2.1**

1. Let $[L]_h$ denotes the $\sim_h$ equivalence class containing language $L$.

2. Then the entropic quotient space $\Omega_\zeta^A$ is the collection $\{[L]_h : L \in \mathcal{P}_A\}$; $L, M, N, \ldots$ will denote points of $\Omega_\zeta^A$, and $L, M, N, \ldots$ will denote subsets of $\Omega_\zeta^A$.

3. The mapping $\eta_h : \mathcal{P}_A \to \Omega_\zeta^A$ takes a language to its $\sim_h$ equivalence class.

**Definition 4.2.2** For any two points, $L$ and $M$ in the entropic quotient space $\Omega_\zeta^A$, the distance $d_h(L, M)$ is defined as follows:

$$d_h(L, M) = \inf \{d_h(L, M) : L \in L, M \in M\}.$$
Since all proofs in subsection 3.3.2 rely on the definition of language pseudo-metrics and norms and a similar definition of the Besicovitch quotient space only, we may recopy the main conclusion here, with application to \( d_h \).

**Lemma 4.4** The mapping \( \eta_h \) is an isometry, and \( d_h(\eta_h(L), \eta_h(M)) = d_h(L, M) \) for all languages \( L, M \in PA \).

From this, the \( d_h \) metric topology of \( QA_h \), which we will denote \( \tilde{\tau}_h \), is the quotient of the \( d_h \) pseudo-metric topology on \( PA \). We will denote the pseudo-metric topology on a language space \( \tau_h \), and call this the entropic language topology. Moreover, a relationship based on the entropic language norm of languages in points of \( QA_h \) is well-defined.

**Lemma 4.5** The relation \( \equiv_h \) defined by \( L \equiv_h M \) if \( L \in L \) and \( M \in M \) implies \( \|L\|_h = \|M\|_h \) is an equivalence relation on \( QA_h \).

**Proof.** Trivial. \( \square \)

**Definition 4.2.3**

1. Let \( \langle L \rangle_h \) denote the \( \equiv_h \) equivalence class of point \( L \) in \( QA_h \).

2. Let \( N_h \) denote space made up of the collection of all \( \equiv_h \) equivalence classes, where \( L, M, N, \ldots \) denote elements of this space and \( L, M, N, \ldots \) denote subsets of this space.

3. Let \( \kappa : QA_h \rightarrow N_h \) map points in \( QA_h \) to their \( \equiv_h \) equivalence classes.

   If \( r \in [0, 1] \), we will denote by \( r_h \) the set \( \{ L \in QA_h : \|L\|_h = r, \text{ if } L \in L \} \) of all points containing languages with norm \( r \).

As with elements of \( N_\zeta \), it is clear that, if \( L \in N_h \), then \( L = r_h \) for some \( r \in [0, 1] \).

Although it is trivial that \( 0_h \) is a singleton in \( QA_h \), the following example illustrates the dissimilarity between \( 1_h \) and \( 1_\zeta \).

**Example 12** Let \( L' \) be an \( r \)-expansive language, where \( r = \frac{2}{3} \). Let \( L'' \subseteq L' \) be an \( r \)-expansive language, where \( r = \frac{1}{3} \). Construct languages \( L_1 \) and \( L_2 \) as follows. Let

\[
L_1^{[k]} = \begin{cases} 
L'^{[k]}, & \text{if } k \text{ is not a power of } 2, \text{ and} \\
A^k, & \text{if } k = 2^i, i \in \mathbb{N}
\end{cases}
\]
and let

\[ \mathcal{L}_2^k = \begin{cases} \mathcal{L}^n[k], & \text{if } k \text{ is not a power of } 2, \\
A^k, & \text{if } k = 2^i, \ i \in \mathbb{N} \end{cases} \]

Then, by Remark 11, \( \| \mathcal{L}_1 \|_h = \| \mathcal{L}_2 \|_h = 1 \), but, first, \( \| \mathcal{L}^c_1 \|_h = 1 \), \( \| \mathcal{L}^c_2 \|_h = 1 \), and, second, \( d_h(\mathcal{L}_1, \mathcal{L}_2) = \frac{2}{3} \).

This is completely at odds with the relationship of elements of \( 1_\zeta \) to each other: elements of \( 1_h \) have their own complements in \( 1_h \); pairs of points in \( 1_h \) need not be closer than \( \frac{2}{3} \) to each other, in \( d_h \) distance; and, thus, that \( 1_h \) contains representatives of different \( \sim_h \) equivalence classes.

Finally, we define distance on \( \mathcal{N}_h \) as follows.

**Definition 4.2.4** Let the function \( \rho : \mathcal{N}_h \times \mathcal{N}_h \to [0, 1] \) map the pair of elements \( \mathcal{L}, \mathcal{M} \in \mathcal{N}_h \), where \( \mathcal{L} = \mathcal{r}_h \) and \( \mathcal{M} = \mathcal{s}_h \), and \( r, s, \in [0, 1] \), to the value \( |r - s| \).

It is obvious that \( \rho \) is a metric. Were there an analog of Proposition 2 in \( \Omega^A_h \), then \( \rho \) would be the quotient of the entropic distance on the upper quotient space and imply, similarly to the conclusion of Theorem 3.1, that \( \mathcal{N}_h \), like \( \mathcal{N}_\zeta \), is homeomorphic to the unit interval, and prove that \( \mathcal{N}_h \) and \( \mathcal{N}_\zeta \) are topologically identical. This, however, is not the case.

4.2.2 A point in the quotient space

From Definition 4.2.1 and Remark 12, all of the factorial languages in a point in the entropic quotient space have the same entropy. We can also draw certain conclusions about the number of words in a \( \sim_h \)-equivalent languages.

**Lemma 4.6** If point \( \mathcal{L} \in \Omega^A_h \), then languages \( \mathcal{L}, \mathcal{M} \in \mathcal{L} \) if and only if, for all \( \epsilon > 0 \), there exists \( N_\epsilon \in \mathbb{N} \) such that \( k > N_\epsilon \) implies \( \#(\mathcal{L} \triangle^k \mathcal{M}) < \alpha^{k\epsilon} \).

**Proof.** Necessity follows from Definition 4.1.1. But so does sufficiency: if there exists \( \epsilon \) such that, for infinitely many \( k \), the cardinality of \( \mathcal{L} \triangle^k \mathcal{M} \) is no less than \( \alpha^{k\epsilon} \), then, for every such word-length \( k \),

\[ \frac{1}{k} \log_{\alpha} \#(\mathcal{L} \triangle^k \mathcal{M}) > \frac{1}{k} \log_{\alpha} \alpha^{k\epsilon} = \epsilon, \]

100
and \(d_h (L, M) > \varepsilon > 0\). Therefore, \(L \not\sim_h M\), and the languages \(L\) and \(M\) cannot be in the same point in \(Q_h^A\).

If we take \(\varepsilon = 1/m\), where \(m \to \infty\), then this gives the following corollary.

**Corollary 4.6.1** \(L, M \in L\) if and only if, for all \(m \in \mathbb{N}, m > 0\), there exists \(k'_m\) such that \(k > k'_m\) implies \(# (L \triangle^k M) < \alpha^{k/m}\).

Consequently, for every language pair \(L, M \in L\), there exists a sequence \(K_h (L, M) = \{k'_m\}_{m \in \mathbb{N}}\), where

\[k'_m = \min \left\{ i : k > i \text{ implies } # (L \triangle^k M) < \alpha^{k/m} \right\}.

Note the distinction between \(K_h\) and the sequence \(K_\zeta\) which characterizes languages in the same point of \(Q_\zeta^A\) (Lemma 3.6 on page 64). Since, in general, \(k/m \ll k - m\), it is reasonable to suppose that a point of \(Q_h^A\) is a far more exclusive set of languages. In fact, languages belonging to a point in \(Q_h^A\) belong to a single point of \(Q_\zeta^A\).

**Corollary 4.6.2** If \(L \sim_h M\), then \(L \sim_\zeta M\).

**Proof.** We claim that languages which fulfill the hypotheses of 4.6.1 also fulfill the hypotheses of Lemma 3.12 on page 78. Let languages \(L\) and \(M\) be such that \(L \sim_h M\). Then there exists \(k'_m \in K_h\) for all \(m > 2\) such that \(k > k'_m\) implies \(# (L \triangle^k M) < \alpha^{k/m}\). This means that if \(k = \max \{k'_m, m + 2\} + 1\), then

\[
k > m + 2 = \frac{m^2 - 1}{m - 1} + 1 > \frac{m^2}{m - 1}, \text{ and so } \quad k(m - 1) > m^2, \text{ i.e., } k - km < -m^2, \text{ or } k < km - m^2.
\]

Consequently, \(k - m > k/m\), and we have that \(# (L \triangle^k M) < k/m < k - m\). This is not enough, because Lemma 3.12 requires that \(# (L \triangle^{<k} M)\) be less than \(# A^{k-m}\). But, since \(k - m > k/m\) so

\[
k - m - k/m \geq 1 \geq \alpha^{k-m} - \# (L \triangle^{<k} M).
\]
The number of words of lengths shorter than $k$ is, at the most, $\frac{\alpha^k - 1}{\alpha - 1}$. Let $N = k + \frac{\alpha^k - 1}{\alpha - 1} + 1$, then consider $l > N$. Since condition (4.5) holds at every word length between $k$ and $l$, we see that

$$\#(L \triangle M) < \frac{\alpha^k - 1}{\alpha - 1} + \sum_{i=k}^{l} (\alpha^{i-m} - 1) < \sum_{i=k}^{l} \alpha^{i-m} < \alpha^{l-m}.$$

This establishes our claim. □

In other words, for each $L_h \in Q_h$, there exists $M_\zeta \in Q_\zeta^A$ such that, if language $L$ is such that $L \in L_h$, then $L \in M_\zeta$. This leads to the following corollary.

**Corollary 4.6.3** *The entropic topology is a refinement of the Besicovitch topology.*

**Proof.** By (4.6), if $d_h(L, M) < \varepsilon$, then $d_\zeta(L, M) < \varepsilon$. Hence, for each basis element $L_h$ of $\tilde{\tau}_h$, there is a basis element $L_\zeta$ of $\tilde{\tau}_\zeta$ such that every language in every point in $L_h$ is contained in some point in $L_\zeta$. □

**Corollary 4.6.4** *A language space is not compact under the entropic topology.*

**Proof.** If every open cover of $\mathcal{P}_A$ which is a subset of $\tau_h$ has a finite subcover, then, from an open cover of $\mathcal{P}_A$ which is a subset of the Besicovitch topology, we can construct an finite subcover, as well. This is impossible, by Corollary 3.7.2 on page 68. Consequently, the entropy topology is not compact, either. □

Note that this leaves open the possibility that the two topologies are equivalent. As it will turn out, when we examine the locally testable languages, this is not the case. In the meantime, we return to the consideration of the sequence $K_h(L, M)$.

**Lemma 4.7** *The symmetric set-difference $L \triangle M$ is finite if and only if in $K_h(L, M)$ there exists $N \in \mathbb{N}$ such that $k_m = k_{m+1}$ if $m > N$.**

**Proof.** Sufficiency is obvious.

If there were no such $N$ that $k_m = k_{m+1}$ if $m > N$, then for infinitely many $k$, $\#(L \triangle^k M) > \alpha^{k/m} > 1$, for some $m > 0$, so $L \triangle M$ is infinite, which proves necessity. □
4.2.3 Elements of the upper quotient space

The geometry induced by the entropic distance is fundamentally different from that of the Besicovitch topology. The entropic topology $\tau_h$ is not, as might first be imagined, a “scaling” of $\tau_\zeta$. Distances, the location of antipodes, and the “shape” of the quotient spaces with respect to the upper quotient space are all widely divergent. We uncover here the source of these distinctions, namely that the image of the entropic topology in the upper quotient space is the discrete topology, except that the set $\{0\}$ is closed, but not open and the interval $(0,1]$ is open, but not closed. In Theorem 4.1, it is shown that this comes about, in part, because every language in every point not in $1_h$, that is, every element of the upper entropic quotient space other than $1_h$, is in the element $0_\zeta$ in the Besicovitch upper quotient space, which is to say, in the point $\eta_\zeta(\Lambda)$.

**Lemma 4.8** If $L \in r_h$ and $M \in s_h$, then $d_h(L, M) \leq \max \{r, s\}$. If $r \neq s$, $d_h(L, M) = \max \{r, s\}$.

**Proof.** Assume languages $L$ and $M$ meet the hypothesized conditions. For each $\varepsilon > 0$, there are at most finitely many integers $k \geq 0$ such that $\#L[k] > \alpha^k(r+\varepsilon)$ or $\#M[k] > \alpha^k(s+\varepsilon)$, so let $n_\varepsilon \in \mathbb{N}$ be such that, if $k > n_\varepsilon$, then neither of these inequalities hold. Then it is clear that, for any sequence $K = \{k_i\}_{i \in \mathbb{N}} \subseteq \mathbb{N}$, such that $k_{i+1} > k_i$ and $k_0 > n_\varepsilon$, $\#(L \triangle M)$ is bound above by the sum $\alpha^{k_i(r+\varepsilon)} + \alpha^{k_i(s+\varepsilon)}$, and, if we assume $r > s$ and choose $\varepsilon < \frac{r-s}{2}$, there is at least sequence $K$ such that, for $i \in \mathbb{N}$, $\#(L \triangle M)$ is bound below by the difference $\alpha^{k_i(r-\varepsilon)} - \alpha^{k_i(s+\varepsilon)}$. If $r = s$, the lower bound is not greater than 0, since $d_h(L, L) = 0$.

Then calculation gives

$$d_h(L, M) = \|L \triangle M\|_h = \limsup_{k \to \infty} \frac{1}{k} \log_\alpha \#(L \triangle^k M)$$

$$< \lim_{i \to \infty} \frac{1}{k_i} \log_\alpha \left( \alpha^{k_i(r+\varepsilon)} + \alpha^{k_i(s+\varepsilon)} \right)$$

$$= \lim_{i \to \infty} \frac{1}{k_i} \log_\alpha \alpha^{k_i(r+\varepsilon)} \left( 1 + \frac{\alpha^{k_i(s+\varepsilon)}}{\alpha^{k_i(r+\varepsilon)}} \right)$$

$$= r + \varepsilon + \lim_{i \to \infty} \frac{1}{k_i} \left[ 1 + \left( \frac{\alpha^s}{\alpha^r} \right)^{k_i} \right]$$

$$= r + \varepsilon,$$

(4.7)

(4.8)
since $\alpha^s / \alpha^r < 1$, by assumption. This proves the first claim of the lemma. If $r > s$, then further calculation shows

$$
\limsup_{k \to \infty} \frac{1}{k} \log_{\alpha} \# \left( L \triangle^k M \right) \geq \limsup_{k \to \infty} \frac{1}{k} \log_{\alpha} \left( \# L^{[k]} - \# M^{[k]} \right)
\geq \limsup_{k \to \infty} \frac{1}{k} \log_{\alpha} \left( \alpha^{k_i (r-\varepsilon)} - \alpha^{k_i (s+\varepsilon)} \right)
= \limsup_{k \to \infty} \frac{1}{k} \log_{\alpha} \alpha^{k_j (r-\varepsilon)} \left( 1 - \alpha^{k_i (s-r+2\varepsilon)} \right)
= \limsup_{k \to \infty} \frac{1}{k} \log_{\alpha} \alpha^{k_j (r-\varepsilon)} \left[ 1 - \left( \frac{\alpha^{s+2\varepsilon}}{\alpha^r} \right)^{k_i} \right],
$$

so that

$$
d_h (L, M) \geq r - \varepsilon + \limsup_{i \to \infty} \frac{1}{k_i} \log_{\alpha} \left[ 1 - \left( \frac{\alpha^s}{\alpha^r} \right)^{k_i} \alpha^{2\varepsilon k_i} \right],
\quad \Rightarrow \quad d_h (L, M) = r - \varepsilon \tag{4.9}
$$

since, by assumption $\alpha^{s+2\varepsilon} / \alpha^r < 1$. As $\varepsilon$ goes to 0, the upper bound established above and (4.9), taken together, prove that, if $r \neq s$, then $d_h (L, M) = \max \{ r, s \}$. \hfill \box

**Corollary 4.8.1** For all $r < 1$, if language $L \in r_h$, then $L^c \in 1_h$.

**Proof.** Although this can also be demonstrated by manipulating the definition, it is a direct consequence of Lemma 4.1 and Lemma 4.8. \hfill \box

**Corollary 4.8.2** For every point $L$ in $Q_A^h$ there is a unique (entropic) antipode, $L^c$, such that $L \in L$ if and only if $L^c \in L^c$.

**Proof.** The proof of Corollary 3.9.1 applies verbatim. \hfill \box

**Corollary 4.8.3** Given language $L \in P_A$ such that $\| L \|_h = r$, and $\varepsilon > 0$, let $B_{\varepsilon}^h (L)$ denote the open neighborhood \{ $M \in P_A : d_h (L, M) < \varepsilon$ \} in $(P_A, \tau_h)$. Then, if language $M \in B_{\varepsilon}^h (L)$,

1. If $\varepsilon < r$, then $M \in r_h$.
2. If \( \varepsilon \geq r \), then \( \|M\|_h < \varepsilon \).

**Proof.** The first conclusion follows from Lemma 4.8: if \( d_h(L, M) < \varepsilon < r \), then \( d_h(L, M) = \max \{\|L\|_h, \|M\|_h\} > \varepsilon \), if \( \|L\|_h \neq \|M\|_h \), which is a contradiction. The second follows similarly: if \( \|M\|_h > \varepsilon \geq r \), then \( d_h(L, M) = \max \{\|L\|_h, \|M\|_h\} \geq \varepsilon \), meaning that \( M \notin B^h_{\varepsilon} (L) \). □

Note that Corollary 4.8.3 does not say precisely what happens within the element \( r_h \). The converse of part 2 of the corollary is obvious: open neighborhoods of radius greater than the norm of the language around which they are formed consist of the collection of all languages of norm less than the radius. But we now inquire into the content of open neighborhoods with radius less than or equal to the norm of the center language. From part 1 of the corollary we know that, however else they may be described, they are contained within the \( \equiv_h \)-equivalence class to which the center belongs.

**Corollary 4.8.4** If \( 1 \geq r > r' \geq 0 \) or \( 1 > r \geq r' \geq 0 \) and the point \( L \in \Omega_h^A \) is such that \( L \in r_h \), then there exists a point \( M \in r_h \) such that \( d_h(L, M) = r' \).

**Proof.** If \( r' = 0 \), this is trivially satisfied by the point \( L \) itself, so assume \( r' > 0 \). There are two cases. Either \( r = 1 \) or \( r < 1 \).

If \( r < 1 \), then, for \( \varepsilon < 1 - r \) and \( i \in \mathbb{N} \), there exists \( k_{\varepsilon, r} \in \mathbb{N} \) such that, if \( k > k_{\varepsilon, r} \) we have

\[
\left( \frac{\alpha^{r'}}{\alpha} \right)^k < \left( \frac{\alpha^{r + \varepsilon}}{\alpha} \right)^k < \frac{1}{2},
\]

and, hence,

\[
\left( \frac{\alpha^{r + \varepsilon}}{\alpha} \right)^k + \left( \frac{\alpha^{r'}}{\alpha} \right)^k < \frac{1}{2} + \frac{1}{2} = 1.
\]

This implies the following inequality if \( k > k_{\varepsilon, r} \):

\[
\alpha^{k(r + \varepsilon)} + \alpha^{kr'} < \alpha^k.
\]

There exists a monotonic strictly increasing sequence \( \{k_i\}_{i \in \mathbb{N}} \subseteq \mathbb{N} \), with \( k_0 > k_{\varepsilon, r} \), such that \( \#L[k_i] > \alpha^{k_i(r - \varepsilon)} \) for all \( i \in \mathbb{N} \). There exists language \( L_{r', i} \subseteq A^k \setminus L[k] \) such that \( \#L_{r', i} = \left\lfloor \alpha^{k_i r'} \right\rfloor \), for all \( i \in \mathbb{N} \), as there are enough words to form such a language by (4.11). Let language \( M = L \cup \bigcup_{i \in \mathbb{N}} L_{r', i} \). Then \( \|L \Delta M\|_h = r' \). Thus, \( d_h(L, M) = r' < r \). But this implies that \( M \in r_h \), for otherwise, \( d_h(L, M) = \max \{r, r'\} = r > r' \), by Lemma 4.8, which is a contradiction.
If $r = 1$, then, setting $\varepsilon = \frac{r - r'}{2}$, there exists a monotonic strictly increasing sequence $\{k_i\}_{i \in \mathbb{N}} \subseteq \mathbb{N}$ such that $\#L[k_i] > \alpha^{k_i}(1 - \varepsilon i + 1)$ for all $i \in \mathbb{N}$. It follows, first, that, for all $i \in \mathbb{N}$, there exists a language $M_i \subseteq L[k_i]$ such that $\#M_i = \left\lfloor \alpha^{k_i}(r - r') \right\rfloor$. Second, setting $M = \bigcup_{i \in \mathbb{N}} M_i$, we find that $d_h(L, M) = r' < 1$. Again, $\|M\|_h = r = 1$, since, otherwise, $d_h(L, M) = \max\{r, r'\} = 1 > r'$, also a contradiction. \hfill \Box

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{The quotient spaces of the entropic distance.}
\end{figure}

The entropic quotient space is comparable to a disk resting on a cone, with complementation projecting each language of norm less than one into the element $1_h$. Curvature of the space is a representation of the fact that neighborhoods of small radii lie completely within a single $\equiv_h$-equivalence class. One such small neighborhood is shown around $L$. The portion of the disk not indicated as extending outside the radius of the cone contains all languages not in $0_h$, while every language of entropic norm less than one is in $0_h$. By contrast to small-radii neighborhoods, neighborhoods with radii larger than the norm of the language they are centered “around” contain all languages of norm less than the radius.

These corollaries establish that $\mathcal{N}_h$ is not homeomorphic to the unit interval in the manner required for the construction of a randomness test, since, by Corollaries 4.8.3 and 4.8.4, open sets projected on $[0, 1]$ by the mapping $\kappa$ must include $\{r\}$, for all $r \in (0, 1]$ and every half-open interval $[0, r)$, for $r \geq 0$, but must exclude $\{0\}$. But the discrete topology is not separable on the unit interval.

It may be noted from (4.10) that the proportion of the cardinality of words of any length in any language in any point of $\Omega^A_h$ other than points in $1_h$ to the total number of words of that length over $A$ goes to 0 as a limit. Thus, the work so far may be summed up in the following theorem.

**Theorem 4.1** If language $L \in \mathcal{P}_A$ is in the element $r_h \in \mathcal{N}_h$, where $r < 1$, then, in the Besicovitch upper quotient space, $L \in 0_c$. 

106
Proof. It is shown that we can disregard finitely many word lengths when calculating the Besicovitch norm. Let $L \in r$ where $r < 1$, let $1 - r > \varepsilon > 0$, and set $k_\varepsilon$ equal to

$$\max \left\{ \left\lceil \frac{\log_\alpha \varepsilon \left[ \alpha^{1-(r+\varepsilon)} - 1 \right]}{r + \varepsilon - 1} \right\rceil , n_\varepsilon \right\} ,$$

where $n_\varepsilon$, as above, is such that $k > n_\varepsilon$ implies $\#L[k] < \alpha^{k(r+\varepsilon)}$. We claim that, for $k_\varepsilon < k' < k''$, $k', k'' \in \mathbb{N}$, the following inequality holds:

$$\frac{\sum_{i=k'}^{k''} \#L[i]}{\alpha^{k''}} < \varepsilon.$$

Since need not consider words in $L$ of length less than $k'$ in calculating the Besicovitch norm of $L$, it will follow that $\|L\|_\zeta = 0$, since $\#L \leq k'' > \alpha^{k''}$.

First, we know that

$$\frac{\sum_{i=k'}^{k''} \#L[i]}{\alpha^{k''}} < \sum_{i=k'}^{k''} \frac{\#L[i]}{\alpha^i} < \sum_{i=k'}^{k''} \frac{\alpha^{i(r+\varepsilon)}}{\alpha^i} = \sum_{i=k'}^{k''} \left( \frac{\alpha^{r+\varepsilon}}{\alpha} \right)^i = \left( \frac{\alpha^{r+\varepsilon}}{\alpha} \right)^{k''} \left( \frac{\alpha^{r+\varepsilon}}{\alpha} \right)^{k'} - \left( \frac{\alpha^{r+\varepsilon}}{\alpha} \right)^{k'} \frac{\alpha^{r+\varepsilon} - 1}{\alpha - \alpha^{r+\varepsilon}} \frac{\alpha^{k''} \alpha^{k''-k'} - \alpha^{r+\varepsilon} k''}{\alpha^{r+\varepsilon}},$$
which simplifies as follows:

\[
\frac{\alpha^{k'(r+\varepsilon-1)+k''(r+\varepsilon)}}{\alpha^{k''}} \cdot \frac{\alpha^{r+\varepsilon}}{\alpha - \alpha^{r+\varepsilon}} = \frac{\alpha^{(k'+1)(r+\varepsilon)-k'} - \alpha^{(k''+1)(r+\varepsilon)-k''}}{\alpha - \alpha^{r+\varepsilon}}
\]

\[
= \frac{\alpha^{k'(r+\varepsilon-1)} - \alpha^{k''(r+\varepsilon-1)}}{\alpha^{1-(r+\varepsilon)} - 1}
\]

\[
= \frac{1}{\alpha^{1-(r+\varepsilon)} - 1} \left[ \left( \frac{1}{\alpha^{1-(r+\varepsilon)}} \right)^{k'} - \left( \frac{1}{\alpha^{1-(r+\varepsilon)}} \right)^{k''} \right]
\]

\[
< \frac{1}{\alpha^{1-(r+\varepsilon)} - 1} \left( \frac{1}{\alpha^{1-(r+\varepsilon)}} \right)^{k'}
\]

since \(\alpha^{k'(r+\varepsilon-1)} > 0\). By the definition of \(k_\varepsilon\) and the selection of \(k' > k_\varepsilon\), it follows that \(\|L\|_\zeta < \varepsilon\).

This explains why the entropic topology is a refinement of the Besicovitch topology, a conclusion which may now be made more exact.

**Corollary 4.8.5** *The topology \(\tau_h\) is strictly finer than the topology \(\tau_\zeta\).*

**Proof.** Let \(L\) be an open set contained in \(r_h\), for some \(r \in (0, 1)\). Then by Theorem 4.1 and Lemma 3.11 there is no open set in \((\mathcal{P}_A, \tau_\zeta)\) contained in \(L\). \(\square\)

Finally, the remaining question about finding languages at specific distances within elements of \(N_h\) is when languages at distance 1 may be found within a single \(h\)-equivalence class.

**Corollary 4.8.6** *For languages \(L, M \in \mathcal{P}_A\), both elements of \(r_h \in N_h\), to be such that \(d_h(L, M) = 1\) it is necessary that \(r = 1\) and sufficient that there exists a point \(N \in \Omega_\zeta^A\) (but not \(\eta_\zeta(\Lambda)\)) such that \(L \in N\) and \(M \in N^c\).*

**Proof.** The claim of necessity that \(r = 1\) follows from Corollary 4.8.1. The sufficiency of membership in non-\(\zeta\) antipodes follows from Theorem 4.1 and Corollary 4.6.2. \(\square\)

### 4.3 Entropy and the Chomsky hierarchy

#### 4.3.1 The finite languages

Lemma 3.17 and Corollary 1.1 hold for \((\mathcal{P}_A, \tau_h)\) as for \((\mathcal{P}_A, \tau_\zeta)\), and for the same reasons. We spell out the conclusion for completeness.
LEMMA 4.9 \( FIN_A \subseteq 0_h \).

4.3.2 Locally testable languages

Topological entropy is a well-investigated aspect of language theory. One of its best-known characteristics is that topological entropy is not dependent upon the number of symbols in the alphabet. We show that this means that the entropic norm preserves the distinction between finite and locally testable languages. Consider the set of local languages with a window-length of 1. These are factorial languages, and so the entropic norm of such languages is in proportion to their topological entropy. Not all such languages have entropy 0, so we expect to find the locally testable languages distributed throughout the entropic topological space. Indeed, since the LOC subfamily contains languages which grow exponentially, it contains languages with non-zero entropic norms.

Note that the construction in Corollary 3.18.1 points to conclusions about generally testable languages in the entropic quotient space opposite to those it implies for the Besicovitch quotient space. Suppose language \( L \) is generally testable, with window-length \( n \) and set of permitted factors \( S \subseteq A^n \). The number of words of length \( k = nq + n - 1 \), as shown in Corollary 3.18.1, is \( s^q \alpha^{n-1} \), and we have

\[
\frac{1}{k} \log_\alpha s^q \alpha^{n-1} = \frac{q}{k} \log_\alpha s + \frac{n-1}{k} \tag{4.12}
\]

But \( q = \lfloor k/n \rfloor \). It follows that \( \|L\|_h = (\log_\alpha s)/n > 0 \). Thus \( \|\cdot\|_h \) preserves the distinction between finite languages and infinite locally testable languages. This observation leads to the following lemma.

LEMMA 4.10 Generally testable languages are dense in \( \mathcal{N}_h \).

Proof. We claim that there is a generally testable language with an entropic norm arbitrarily close to that of any language in \( \mathcal{P}_A \). Suppose that \( L \in \mathcal{P}_A \), and that \( \|L\|_h = r \). Let \( \varepsilon > 0 \). If \( \varepsilon \geq \min \{r, 1 - r\} \), then any language in either \( 0_h \) or \( 1_h \) will satisfy the claim, so assume that \( (r - \varepsilon, r + \varepsilon) \subseteq (0, 1) \). Then there exists \( n \in \mathbb{N} \) such that \( n \varepsilon > \log_\alpha 2 \). Since \( \alpha^{(r-\varepsilon)n} = \alpha^{mn}/\alpha^{n\varepsilon} > 1 \) (because, by our assumption, \( r > \varepsilon \)), we have that there exists \( s \in \mathbb{N} \) such that

\[
\alpha^{(r-\varepsilon)n} < s < 4\alpha^{(r-\varepsilon)n} \leq \alpha^{2\log_\alpha 2 \alpha^{(r-\varepsilon)n}} < \alpha^{2n\varepsilon \alpha^{(r-\varepsilon)n}} = \alpha^{(r+\varepsilon)n}.
\]

109
It follows that
\[ r - \varepsilon < \frac{1}{n} \log_{10} s, \]
and by comparison to equation (4.12) and the deduction following the equation, we see that if we select \( s \) permitted factors of length \( n \), we have the required generally testable language.

From this it follows that \( \text{REG}_A \) is dense in \( \mathcal{N}_h \) as are the context-sensitive, context-free, and recursively enumerable languages. But it shows the existence, as well, of a subfamily of the regular languages members of which can be separated by open sets in \( \mathcal{Q}_h^A \), but not in \( \mathcal{Q}_\zeta^A \). This proves the following corollary, obvious, in any case, from the near-complete discreteness of the topology.

4.4 Randomness in the entropy topology

Although there are constructions available on the entropic topological language space that are formally similar to those on the Besicovitch space, the lack of a homeomorphism between \( \mathcal{N}_h \) and \([0, 1]\) means that the randomness space we will discuss here has meaning only on the upper quotient space, \( \mathcal{N}_h \). In particular, if the randomness conclusions are applied to the language space, every language outside the element \( 0_\zeta \) of \( \mathcal{N}_\zeta \) would be declared nonrandom, together with uncountably languages found in the zero point of the Besicovitch quotient space. At the same time, every random language on the entropic randomness space would necessarily be declared nonrandom in the Besicovitch randomness space. The following subsections are numbered so as to match the enumeration in Subsection 1.5.2.

4.4.1 Separability

The upper quotient space, regarded as a homeomorphic image of the unit interval under the metric \( \rho \), is separable. This is the strategy of Chapter Three, but on the entropic topology, this homeomorphism has no connection to open sets in the language space, as observed above. Nevertheless, we will consider the consequences of taking randomness in this sense to define the randomness of languages.
4.4.2 Measure

The unit interval, regarded as a randomness space under Lebesgue measure \( \mu_u \) has an enumerable basis \( B_u \) if we use intervals with end-points of the form \( n\alpha^{-m} \) and \((n+1)\alpha^{-m}\). The triple \( ([0,1], B_u, \mu_u) \) is a randomness space. Every open set in the metric space \( (\mathcal{N}_h, \rho) \) has an open image in \([0,1]\) and is therefore \( B_u \)-computable if it is the union of sets found in the \( \emptyset \), so this, while we remain on the upper entropic quotient space, so metrized, we can directly adapt all conclusions from the Besicovitch upper quotient space to the entropic upper quotient space.

4.4.3 Randomness tests

Randomness tests \( \mathcal{U} = \{U_i\} \) now have the same definition as on the unit interval: \( \mu_u(U_i) < 2^{-i} \), where there is a recursively enumerable sequence of integers \( N \) such that

\[
U_i = \bigcup_{j \in N} B_{uj}.
\]

Since 1 is constructible, if we define nonrandom languages as languages belonging to nonrandom elements of \( (\mathcal{N}_h, \rho) \), then all languages not belonging to \( 0_\zeta \) are nonrandom, by Theorem 4.1. Moreover, non-constructible real numbers in the unit interval are the image under \( \kappa \) of certain languages belonging to \( 0_\zeta \). All languages in \( 0_\zeta \) are nonrandom, in the Besicovitch randomness space.

4.4.4 Random languages

Nonrandom elements of \( \mathcal{N}_h \) are those which are elements of the intersection \( \bigcap_{i \in N} U_i \) where \( \{U_i\}_{i \in N} \) is a randomness test. If, despite the counter-intuitive results mentioned above, we define randomness on \( (\mathcal{P}_A, \tau_h) \) by this means, the most significant effect it has is to discern nonrandom languages in \( 0_\zeta \). Although the distinction between this forced definition of and the two previous forms of randomness is clear, and we have both that some nonrandom languages in the entropic randomness space are random under the Besicovitch topology, and vice versa, this procedure raises further questions.

Is there a correct way to define randomness in the entropic topology? It is apparent that if we disregard what it says about \( 1_h \), the remaining space can be looked upon as having taken a magnifying

111
glass to a single point in the Besicovitch space, namely $\Omega_\zeta$. By L-rotation, we can perhaps apply the same inspection to every point in $\Omega^A_\zeta$. Then we would have that these pseudo-metric complement each other, not simply in a geometrical sense, but in their ability to gauge language complexity. Randomness may be best discovered in the immediate surroundings by the entropic topology, whereas randomness may be best disclosed on the language space as a whole by the Besicovitch topology. At the same time, each topology, we find that entire equivalence classes are composed of random, non-RE languages.
The aim of this chapter is to make use of the results from the investigation of language pseudo-metrics, both specific and general, to gain some insight into the transformations of one language into another, i.e., language morphisms. We preface the main discussion with a summary of the structures developed in Chapters Two, Three, and Four. In general, we wish to know what properties of languages are preserved by which language morphisms and under which topologies. For the most general conclusions we consider a language morphism mapping languages from one language space into a distinct language space, topologized differently. Other conclusions are restricted to mappings between language spaces under the same topology; still others require that the morphism be from a language space to itself.

Since the language pseudo-metrics inducing these topologies are based on the symmetric set-difference or permutative set-difference, both of which can be partially ordered by set-inclusion, we also desire a general formulation of the action of a morphism with respect to word-length. Such observations are independent of the pseudo-metric. Of interest are both the distribution of words of constant length under the morphism and the pre-image of words of constant length under the morphism.

We will assume that we are dealing with language pseudo-metrics, and perhaps metrics. Then, following the usage in previous chapters, each pseudo-metric is denoted with a distinct subscript. In the general case, we will use the symbols $d_\nu, d_\xi, d_\psi$, etc., for unspecified language pseudo-metrics and say

1. that $d_\nu$ is an arbitrary language pseudo-metric with a corresponding language norm $\| \cdot \|_\nu$,

2. that $\Omega_\nu^A$ is the quotient space defined by $\sim_\nu$, where $L \sim_\nu M$ if $d_\nu(L, M)$, and points in $\Omega_\nu^A$ are denoted in sans serif font
3. that $\eta_\nu$ denotes the quotient map from $\mathcal{P}_A$ into $\Omega^A_\nu$ under which the quotient metric $d_\nu$ is an isometry of $d_\nu$ and

4. the quotient topology $\tilde{\tau}_\nu$ is the image under $\eta_\nu$ of the pseudo-metric topology $\tau_\nu$ induced by $d_\nu$.

5. Whenever necessary, $r_\nu$ will denote the class of points in the quotient space containing all languages at distance $r$ from the point $\Lambda$, which is to say, all languages of norm $r$.

Where possible, of course, $\mathcal{N}_\nu$ will be the unit interval. This requires not only some type of normalization of the norm, but its surjectivity on $[0,1]$, so we do not assume $\mathcal{N}_\nu \cong [0,1]$ in this chapter. It was said in Section 1.3, that semiring morphisms preserve aspects of language structure. We proceed from this standpoint.

**Definition 5.0.1** If $A$ and $B$ are alphabets,

1. let $\mathcal{M}(A,B)$ denote the collection of language morphisms from $\mathcal{P}_A$ into $\mathcal{P}_B$, a subset of the larger class of monoid morphisms;

2. let the collection of language space codes from $\mathcal{P}_A$ to $\mathcal{P}_B$ will be denoted $\mathcal{C}(A,B)$; and

3. let $\mathcal{W}(A,B)$ denote the collection of non-expansive, non-erasing language morphisms from $\mathcal{P}_A$ into $\mathcal{P}_B$ — which are necessarily injective, — i.e., every language morphism $\Phi_{\varphi}$ which is an extension (see Proposition 1) of an arbitrary word morphism $\varphi : A^* \rightarrow B^*$.

It is clear from earlier remarks that $\mathcal{M}(A,B) \supseteq \mathcal{W}(A,B)$ and $\mathcal{C}(A,B) \supseteq \mathcal{W}(A,B)$.

5.1 The continuity of morphisms

The continuity morphism $\Phi \in \mathcal{M}(A,B)$, one for which every open set of languages in $\mathcal{P}_B$ is the image under $\Phi$ of an open set in $\mathcal{P}_A$, would at first seem to depend on the nature of the morphism and the character of the language pseudo-metrics used to topologize the two spaces. We show that dependence of continuity upon the language pseudo-metric takes the form of dependence upon the number of symbols in the alphabets of the spaces.

The first question, then, is “Which morphisms are continuous in a language space under topologized by a pseudo-metric $d_\nu$?” Continuity between two language spaces topologized by $d_1$ is a
straight-forward, since $d_1$ is a metric. Language morphism $\Phi : \mathcal{P}_A \to \mathcal{P}_B$ is continuous if and only if every language $M$ in every cylinder set $C_{G_M,m} \subseteq \mathcal{P}_B$ is in the image of a some language $L$ in a cylinder set $C_{F_L,k} \subseteq \mathcal{P}_A$, i.e., $\Phi (C_{F_L,k}) \subseteq C_{G_M,m}$. That is, it must be that, for each $m \in \mathbb{N}$ and each finite language $G_M \subseteq B^<m$, there exists $k \in \mathbb{N}$ and language $F_L \subseteq A^<k$ such that $\Phi (L) \cap B^<m = G_M$ if and only if $L \cap A^<k = F_L$. This requires the surjectivity of $\Phi$. Because $\Phi$ is a language morphism, $\Phi (F_K) = G_M$. But then $\Phi$ induces a bijection between $\text{FIN}_A$ and $\text{FIN}_B$, and so, by Fact 1 on page 12, $\Phi$ is injective. Consequently, by 1.2, $\Phi$ is a natural isomorphism between $\mathcal{P}_A$ and $\mathcal{P}_B$, and $\# A = \# B$.

Continuity under a pseudo-metric topology $\tau_\nu$ is synonymous with continuity on the metric quotient space induced by the equivalence relation $\sim_\nu$. This is not inconsistent with the situation in metric topologies. For example, adopting our notation on the Cantor space, we could define “$L \sim_1 M$” to mean $L = M$, so that equality of points and openness of sets on the quotient space $\Omega^A_\nu$ is synonymous with equality of languages and openness of sets in $\mathcal{P}_A$. The difference between this and a strictly pseudo-metric topology is that the points in $\Omega^A_\nu$ do not, in general, correspond to a set of languages of a particular structure.

**Definition 5.1.1** The language mapping $\Phi : (\mathcal{P}_A, \tau_\nu) \to (\mathcal{P}_B, \tau_\xi)$ is continuous if and only if $\Phi$ is continuous as a map from $\Omega^A_\nu$ into $\Omega^B_\xi$.

Thus, under a continuous language morphism, entire $\sim_\nu$ equivalence classes are mapped to sets of $\sim_\xi$ classes. The following are other observations regarding surjectivity.

**Lemma 5.1** If language mapping $\Phi : (\mathcal{P}_A, \tau_\nu) \to (\mathcal{P}_B, \tau_\xi)$ is continuous, then $\Phi$ is surjective on $\Omega^B_\xi$.

**Proof.** Suppose $\Phi$ is not surjective. Then $\Omega^B_\xi$ is open under the $\tau_\xi$ topology yet there is no open set $L \subseteq \Omega^A_\nu$ such that $\Phi (L) \supseteq \Omega^B_\xi$. Therefore, $\Phi$ is not continuous. □

**Corollary 5.1.1** If $\# B > \# A$, no element of $\mathcal{W}(A, B)$ is continuous.

**Corollary 5.1.2** Natural language isomorphisms are continuous between language spaces topologized under the same topology.
On the other hand, the injectivity of a language morphism is clearly not required for continuity. An example of this is the following morphism, a quotient map between two spaces.

**Example 13** Consider the morphism $\Phi \in \mathcal{M}(\mathbb{N}, \mathbb{N})$ which is an extension of the word morphism $\varphi$ such that

$$
\begin{align*}
\varphi(1) &= \varphi(3) = 1 \\
\varphi(2) &= \varphi(4) = 2,
\end{align*}
$$

and consider the language $L = \{1\} \in \mathcal{P}_2$. The languages $\{1\}, \{3\}$, and $\{1, 3\}$ comprise the inverse image of $L$. A cylinder of length $n$ set around $\{1\} \in \mathcal{P}_2$ is the image, under $\Phi$, of three cylinder sets in $\mathcal{P}_4$, namely $C_{\{1\}, n}, C_{\{3\}, n}$, and $C_{\{1, 3\}, n}$, i.e., one around each of these languages. This clearly can be extended to any language in $\mathcal{P}_2$.

A rather restricted generalization regarding continuity of language codes is presented.

**Lemma 5.2** A language code $\Phi : (\mathcal{P}_A, \tau_\nu) \to (\mathcal{P}_B, \tau_\xi)$ is continuous on its image if $\tau_\nu$ is coarser than $\tau_\xi$.

**Proof.** Let language $L$ be in $\mathcal{P}_A$ and suppose morphism $\Phi \in \mathcal{C}(A, B) \cap \mathcal{M}(A, B)$. If we suppose, toward contradiction, that $\Phi$ is not continuous on $\Phi(\mathcal{P}_A)$, then there is a basis element, $M = B_\xi(\Phi(L)) \in \tau_\xi$ such that, for every $\delta > 0$ there is a language $L' \in \mathcal{P}_A$ such that $d_\nu(L, L') < \delta$ but $d_\xi(\Phi(L), \Phi(L')) > \varepsilon$. Since $d_\nu$ and $d_\xi$ are language pseudo-metrics, $d_\nu(L, L') = ||L \triangle L'||_\nu$ and $d_\xi(\Phi(L), \Phi(L')) = ||\Phi(L) \triangle \Phi(L)||_\xi$. But $\Phi(L) \triangle \Phi(L') = \Phi(L \triangle L')$. Thus our assumption is that there is a there is a positive lower bound to the norm $||\Phi(N)||_\xi$ as the norm of $N$, i.e., $||N||_\nu$, goes to zero. But $\tau_\nu$ is coarser than $\tau_\xi$, meaning that an every open set around some element of $0_\xi$ contains an open set around some element of $0_\nu$. This is a contradiction, so we conclude that $\Phi$ is continuous.

But even the following proposition, which sounds rather obvious, has (as yet) no confirmation.

**Conjecture 2** Every language code is continuous on its image.

A more general issue is, given a language distance and the topology it induces, where does a morphism map different families of languages? One possibility is a mapping which, although not trivial,
is nilpotent with respect to a language pseudo-metric. We define this property for $d_\nu$ and show how it can be used in the case of $d_\zeta$.

**Definition 5.1.2** A language mapping $\Phi : (\mathcal{P}_A, \tau_\nu) \rightarrow (\mathcal{P}_B, \tau_\zeta)$ from $\mathcal{P}_A$ into $\mathcal{P}_B$ will be called $d_\zeta$-trivial if $\Phi(\mathcal{P}_A) \subseteq 0_\zeta \subseteq Q_\zeta$.

What are the qualifications for a language code to be $d_\zeta$-trivial?

**Lemma 5.3** Every language code $\Phi : (\mathcal{P}_A, \tau_\nu) \rightarrow (\mathcal{P}_B, \tau_\zeta)$ such that $\Phi(\{A\}) \subseteq B^k$ for some $k \in \mathbb{N}$, is $d_\zeta$-trivial.

**Proof.** Let $\#A = \alpha$, as usual, and let $\beta = \#B$. Then, for all languages $L \in \mathcal{P}_A$, the language $\Phi(\{A\})$ is the subset of a generally testable language $M$ with window-size $k$ and the set of permitted factors $\Phi(\{A\})$. But this implies, by Lemma 3.18, that $\Phi(\{A\}) \subseteq 0_\zeta$.

Note that the results in Lemma 3.20 can be obtained via this lemma. For the projection of a language over two symbols into a language space containing more than two symbols is obviously a language code. But such a projection is $d_\zeta$-trivial, by Lemma 5.3, since a set of two symbols is a proper subset of an alphabet over two symbols. Consequently, all languages (including all non-RE languages) over two symbols are mapped into $0_\zeta$ in a language space over an alphabet of cardinality greater than two. On the other hand, every alphabet $A$ containing more than two symbols can be encoded into a set of words over two symbols, of some constant word-length $k$, where $2^k > \alpha$. Thus, every language (including every non-RE language) in a language space $\mathcal{P}_A$ over more than two symbols maps to a language in $0_\zeta$ in a language space over two symbols.

We observe that morphisms on topologized language spaces require much more research. In the following section we address a question independent topology, but of importance in any distance function tied to symmetric set-differences, and therefore to every language pseudo-metric. We need to know where the words, in terms of word-length, the image of a language morphism comes from, and where, likewise in terms of word-length, the words in a language map to under a given language morphism.

### 5.2 The pre-image of a section of words

We confine our consideration to language codes. This allows us to denote the language code $\Phi : \mathcal{P}_A \rightarrow \mathcal{P}_B$ as simply $\varphi$, the code from $A^*$ into $B^*$ of which it is the extension, by Theorem 1.1.
The question we address in subsection one may then be put, where do elements of \( \varphi(A^*)^{[k]} \) come from? Of course, \( \# \left[ \varphi(A^*)^{[k]} \right] = \#\varphi^{-1}(B^k) \), because \( \varphi \) is injective. This seems useful since language pseudo-metrics respect the partial order of a language space under set-inclusion, and, therefore, the partial order of languages by section, i.e., word-length. Conversely, in subsection two, we similarly formalize an approach to the question, where in \( B^* \) do elements of \( A^k \) go under the morphism \( \varphi \)?

5.2.1 The word symbol matrices and symbol-length vectors of morphisms

Some algebraic machinery is proposed for the further study of the word-lengths in the image of a code.

For a given \( k \in \mathbb{N} \), we wish to count the number of words in \( \varphi(A^*) \) which are in \( B^k \). Consider the symbols in \( A \), which may be enumerated \( a_1, a_2, \ldots, a_\alpha \). These are mapped to words \( w_1, w_2, \ldots, w_\alpha \) in \( B^* \). Let \( m = \min \{ |w_i| : i \in \mathbb{N}_\alpha \} \), \( M = \max \{ |w_i| : i \in \mathbb{N}_\alpha \} \), the length of the shortest of and longest of these words, respectively. Then \( \varphi(A) = \bigcup_{i=m}^{M} B_i \), where \( B_i \subset B^i \), a union of possible empty subsets of \( B^m \) through \( B^M \). These induce a partition on \( A \) such that \( A_i \subseteq A \) and \( \varphi(A_i) = B_i \), for \( m \leq i \leq M \). Let \( \alpha_i = \#A_i \), for each set \( A_i \). Let \( \Gamma \) be the increasing sequence of integers \( \Gamma = \{ i : A_i \neq \emptyset \} \), where \( g = \#\Gamma \). Let \( \gamma \) be an ordering of \( \Gamma \) such that \( \gamma_i < \gamma_j \) if \( i < j \).

Thus, \( \Gamma = \{ m = \gamma_1, \gamma_2, \ldots, \gamma_g = M \} \).

The non-empty members of the partition of \( A \) are therefore the sets \( A_{\gamma_1}, \ldots, A_{\gamma_g} \).

For each \( s \in \mathbb{N}_g \), if \( \alpha_s = t \), let the function \( a_s \) order the set \( A_s \), so that it can be listed \( A_s = \{ a_{i1}, a_{i2}, \ldots, a_{it} \} \). Finally, let \( h = \max \{ \alpha_i : i \in \Gamma \} \), and let \( \alpha_\varphi \in \mathbb{N}^{1 \times g} \) be the vector \( (\alpha_{\gamma_1} \cdots \alpha_{\gamma_g}) \).

**Definition 5.2.1** Associate every word \( w \) in \( A^* \) with a \( \varphi \) symbol-matrix, \( W_\varphi^w = [w_{ij}] \in \mathbb{N}^{g \times h} \), where \( w_{ij} = k \) if \( |w|_{a_{ij}} = k \) for \( 0 \leq i < g, 1 \leq j \leq h \), and where \( w_{ij} = 0 \) if \( j > \alpha_i \).

Clearly the \( \varphi \) symbol matrix of a word is unique up to permutations of the symbols of the symbols in the word.

If vector \( u = (u_1, \ldots, u_v) \), then denote by \( \Sigma u \) the sum of the components of vector \( u \), viz. \( u_1 + u_2 + \cdots + u_v \). Denote by \( w_i \) the \( i \)th row-vector of \( W_\varphi^u \).
Given \( w \in A^* \) and \( W^\varphi_w \), the \( \varphi \) symbol-matrix of \( w \), the \( \varphi \) symbol-length vector of \( w \) is the vector \( w^\varphi = (\Sigma w_1, \Sigma w_2, \ldots, \Sigma w_g) \).

Then the following is obviously the case:

\[
|\varphi(w)| = \langle w^\varphi, \alpha^\varphi \rangle,
\]

where \( \langle \cdot, \cdot \rangle \) is the usual dot product of vectors. Conversely, if \( x \) is a \( g \)-component vector over \( \mathbb{N} \), and

\[
\langle x, \alpha^\varphi \rangle = k
\]

then for any word \( x \in A^* \) which has exactly \( x_i \) symbols in \( A_{\gamma_i} \), for \( 1 \leq i \leq g \), then \( |\varphi(x)| = k \).

For word \( w \in A^* \), \( |w| = j \) and \( |\varphi(w)| = k \) if and only if \( |w^\varphi| = j \) and \( \langle w^\varphi, \alpha^\varphi \rangle = k \).

We can therefore determine the number of elements of \( \varphi^{-1}(B^k) \) in three stages. First, we need to know the total number of vectors \( x \) over \( \mathbb{N}^{1 \times g} \) such that \( \Sigma x = l \) and \( \langle x, \alpha^\varphi \rangle = k \), for \( \left\lfloor \frac{k}{M} \right\rfloor \leq l \leq \left\lceil \frac{k}{M} \right\rceil \), which is to say, the number of possible \( \varphi \) symbol-length vectors. Next we require the number of \( \varphi \) symbol-matrices \( W^\varphi_x \), that is, \( g \times h \) matrices over \( \mathbb{N} \) the column vectors of which sum to the entries in a given \( \varphi \) symbol-length vector, with \( w_{ij} = 0 \) if \( j > \alpha_i \). Finally, given a \( \varphi \) symbol matrix \( W^\varphi_x \), we need to calculate the number of distinct words \( w \) such that \( |w|_{a_{ij}} = w_{ij} \) for each symbol \( a_{ij} \in A \), where \( 1 \leq i \leq g, 1 \leq j \leq h \).

The final step is standard: the number of distinct permutations of \( \Sigma w^\varphi = |w| \) symbols of which there are \( w_{ij} \) instances each of symbols \( a_{ij} \in A \), where \( 1 \leq i \leq g, 1 \leq j \leq h \) is \( \pi^\varphi(w) \) where

\[
\pi^\varphi(W^\varphi_w) = \frac{|w|!}{\prod_{1 \leq i \leq g} w_{ij}!}, \quad \text{(5.1)}
\]

which is defined for any matrix \( D \in \mathbb{N}^{g \times h} \), including those where \( D_{ij} = 0 \) if \( j > \alpha_i \). Let \( \Xi_k \) be the set of \( k \)-sum \( \varphi \) symbol-length vectors, \( \Xi_k = \{ x \in \mathbb{N}^g : \langle x, \gamma^\varphi \rangle = k \} \). Then, for each \( x \in \Xi_k \), form the set \( Y_{k,i}(x) \) of possible \( i^{th} \) row vectors of corresponding \( \varphi \) symbol-matrices. That is, let \( Y_{k,i}(x) = \{ y \in \mathbb{N}^h : \Sigma y = x_i, y_j = 0 \text{ if } i > \alpha_{\gamma_i} \} \). Every word in \( \varphi^{-1}(B^k) \) is one of the permutations of a word \( w \) whose \( \varphi \) symbol matrix can be represented \([y_{ij}] \in \mathbb{N}^{g \times h}\) such that there
is an \( x \in \Xi_k \) such that, for each \( i \) from 1 to \( g \), \((y_{i1}, y_{i2}, \ldots, y_{ih}) \in Y_{k,i}(x)\). We have from Lemma 5.4 that

\[
\# \varphi^{-1}(B^k) = \sum_{x \in \Xi_k} \prod_{i=1}^{g} \sum_{y \in Y_{k,i}(x)} \pi_{\varphi}([y_{ij}])
\]  \hspace{1cm} (5.2)

This equation may be interpreted as a commutative regular expression for the general word \( w \) in \( A^* \) such that \( \varphi(w) \in B^k \). Let \( a_1, a_2, \ldots, a_g \) be the formal vectors over \((A \cup \{\lambda\})^h\) such that \( a_i = (a_{i1}, \ldots, a_{ih}, \lambda, \ldots, \lambda) \), and let these be the column vectors of a formal matrix \( A_{\varphi} \), an element of \((A \cup \{\lambda\})^{h \times g}\). For all \( x \in \Xi_k \), and any combination of \( y_1, \ldots, y_g \) such that \( y_i \in Y_{k,i} \), for \( 1 \leq i \leq g \), then \( w \in [d_{ij}]A_{\varphi} \), the formal matrix product of the matrix expressing the number of symbols and the matrix containing the symbols, where “+” represents set union and where “·” is commutative concatenation. Then the formal expression for all elements of \( A^* \) which map under \( \varphi \) to words in \( B^* \) of length \( k \) is

\[
\sum_{c \in C_k} \prod_{i \in \mathbb{N}_g} \sum_{d_i \in D_{c_i}} [d_{ij}]A_{\varphi}
\]  \hspace{1cm} (5.3)

regarded as a commutative regular expression. An example follows.

**Example 14** Let \( \varphi \in \mathcal{C}(A, B) \), where \( A = \{a, b\} \), \( B = \{d, e, f\} \) such that

\[
\begin{align*}
\varphi: a & \mapsto de \\
\varphi: b & \mapsto f
\end{align*}
\]  \hspace{1cm} (5.4) \hspace{1cm} (5.5)

Then \( m_\varphi = 2 \), \( A_1 = \{b\} \), \( A_2 = \{a\} \), \( \Gamma = \{\gamma_1 = 1, \gamma_2 = 2\} \), \( \alpha_1 = \alpha_2 = 1 \), and \( \alpha_\varphi = (1 1) \). If \( k \) is even, i.e., \( k = 2l, l \in \mathbb{N} \), then \( C_k = \{(2i, l - i) \mid 0 \leq i \leq l\} \) and if \( k = 2l + 1 \), then \( C_k = \{(2i + 1, l - i) \mid 0 \leq i \leq l\} \). The vectors in each of these may therefore be indexed by \( i \in \mathbb{N}_l \cup \{0\} \).

For each \( k \in \mathbb{N} \) and each \( c \in C_k \), \( D_{c_1} = \{(c_1)\} \) and \( D_{c_2} = \{(c_2)\} \), while \( |c| = c_1 + c_2 \), which is to say \( |c| = l + l \) if \( k = 2l \) and \( |c| = l + i + l \) if \( k = 2l + 1 \). Thus (5.3) simplifies to (5.1), since there is only one partition of each set \( D_{c_i} \) for \( i \in \mathbb{N}_g = \{1, 2\} \). Further, each vector \( d_i \) has only one component for \( i \in \{1, 2\} \). Thus, for \( k = 2l \), the total number of words in \( A^* \) with \( \varphi \)-images in \( B^k \) is given by

\[
\# \left[ \varphi(A^*) \cap B^k \right] = \sum_{i=0}^{l} \frac{(l + i)!}{(2i)!((l - i)!}
\]
and similarly, if \( k = 2l + 1 \), by

\[
\# \left[ \varphi(A^*) \cap B^k \right] = \sum_{i=0}^{l} \frac{(l + i + 1)!}{(2i + 1)!((k - i)!).}
\]

Calculation shows that, therefore,

\[
\# \left[ \varphi(A^*) \cap B^k \right] + \# \left[ \varphi(A^*) \cap B^{k+2} \right] = \# \left[ \varphi(A^*) \cap B^{k+2} \right] \quad (5.6)
\]

for, if \( k = 2l \),

\[
\sum_{i=0}^{l} \left[ \frac{(l + i)!}{(2i)!(l - i)!} + \frac{(l + i + 1)!}{(2i + 1)!(l - i)!} \right] = 1 + \sum_{i=1}^{l} \left[ \frac{1 + \frac{2i}{l - i + 1}}{(2i)!(l - i)!} \right] + \frac{[(l + 1) + (l + 1)]}{[2(l + 1))! [(l + 1) - (l + 1)]} + \sum_{i=1}^{l} \left[ \frac{l + i + 1}{l - i + 1} \frac{(l + i)!}{(2i)!(l - i)!} \right] + \frac{[l + 1 + 0]}{(2 \cdot 0)!(l + 1) - 0]}
\]

\[
= \sum_{i=0}^{l+1} \frac{(l + 1 + i)!}{(2i)!(l + 1 - i)!} \quad (5.7)
\]

and similarly for \( k = 2l + 1 \). Since obviously \( \varphi(A^*) \cap B = \{ f \} \) and \( \varphi(A^*) \cap B^2 = \{ de, ff \} \) it follows that \( \# \varphi \left( A^* \cap B^k \right) = F_{k+2} \), where \( \{ F_n \}_{n \in \mathbb{N}} \) is the Fibonacci sequence. If \( L_1, L_2 \in \mathcal{S}_A \), then, since \( \varphi(L_1) \) and \( \varphi(L_2) \) are factorial, the following inequality holds:

\[
d_h (\varphi(L_1), \varphi(L_2)) < \log_3 \frac{1 + \sqrt{5}}{2}.
\]

5.2.2 Distribution of the morphic image of a language space

Suppose that language code \( \Phi_\varphi \), which we will continue to denote \( \varphi \), is not surjective. If we accept that it is continuous on its image as conjectured, then where is that image? Let \( A \) be partitioned as above into a the disjoint union of sets of symbols such that, in each set, the \( \varphi \)-image of each symbol is a word of the same length. Let these sets be enumerated \( A_1, A_2, \ldots, A_M \), some of which may be empty, and so that \( \alpha \in A_i \) implies \( |\varphi(\alpha)| = i \). Further, define, as above, \( \alpha_i = \#A_i \), and let the
set $Q = \{q_1, \ldots, q_M\}$ be an alphabet of $M$ symbols. Consider the formal polynomial over $\mathbb{N}[Q]$, $Q(\varphi, A^j) = (\sum_{i=1}^{M} \alpha_i q_i)^j$. Then, if $D = \{d_1, \ldots, d_{n(j)}\}$ is the set of coefficients of the terms of $Q [\varphi, A^j]$, it is clear that $\sum_{i=m}^{M} d_i = \alpha^j = |A^j|$. Another way to express this is to say

$$\sum_{i \in \mathbb{N}} \frac{d_i}{\alpha^j} = 1 \quad (5.8)$$

Examining the terms of this polynomial reveals the following. The sum of the subscripts of the $q$s in a term with a non-zero coefficient gives the length of one word in $\varphi(A^j)$. Let the polynomials $Q_t [\varphi, A^j]$ over $\mathbb{N}[Q]$, for $0 \leq t \leq M$, be the sum of all the terms of $Q [\varphi, A^j]$ of the form

$$\alpha_{i'}^{\beta_{i'}} \cdots \alpha_{i''}^{\beta_{i''}} q_{i'}^{\beta_{i'}} \cdots q_{i''}^{\beta_{i''}},$$

where $0 \leq i' \leq i'' \leq M$, and where $\sum_{s=i'}^{i''} i_s \beta_s = t$, and $\sum_{s=i'}^{i''} \alpha_s \beta_s = t$. Each term of $Q_t [\varphi, A^j]$ can be seen to represent a distinct set of words in $\varphi(A^j) \cap B^t$. Then we have the following.

**Lemma 5.5** The evaluation of the polynomial $Q_t [\varphi, A^j]$, where

$$q_1 = q_2 = \cdots = q_M$$

is

$$Q_t [\varphi, j] (1, 1, \ldots, 1) = \# \left( \Phi \left( A^j \right) \cap B^t \right).$$

That is, the sum of the coefficients of $Q_t [\varphi, j]$ is the cardinality of the set of words of length $t$ in the image of $A^j$.

**Corollary 5.5.1** The sequence $\left\{ \frac{Q_t [\varphi, j] (1, 1, \ldots, 1)}{\alpha^j} \right\}$ is a probability distribution of $\varphi(A^j)$ over $B^*$.  

**Proof.** Obvious, from (5.8). \qed

For any $k \in \mathbb{N}$ and language $L \in \mathcal{P}_A$, define the following polynomial.

**Definition 5.2.3** Let $\varphi$ be a language code, which partitions $A$ into sets $A_1, \ldots, A_M$ such that $a \in A_i$ if $|\varphi(a)| = i$, and where $\#A_i = \alpha_i$. Let $Q [\varphi, L[k]]$ denote the $\varphi, k$ polynomial of $L$, where

$$Q [\varphi, L[k]] = \sum_{w \in L[k]} Q(w)$$
and, for $|w| = j$, $Q(w)$ is the monomial $\prod_{i=1}^{j} q(w[i])$, and $q(a) = \alpha_l q_l$ if $a \in A_l$.

The result is that, in the $\varphi, k$ polynomial of $L$, $Q[\varphi, L[k]]$, we find the term $d q_{\beta_1}^\alpha q_{\beta_2}^\delta \cdots q_{\beta_v}^\delta$ if there are exactly $d$ words which are products of the images of $\beta_1$ words in $A_{\delta_1}$, $\beta_2$ words in $A_{\delta_2}$, and so on, up to $\beta_v$ words in $A_{\delta_v}$.

**Definition 5.2.4** Given $x, y \in \mathbb{N}[q_1, \ldots, q_n]$, so that $x = x_1 q_1 + \cdots + x_n q_n$ and $y = y_1 q_1 + \cdots + y_n q_n$, then the partial order $\preceq \subseteq \mathbb{N}[q_1, \ldots, q_n] \times \mathbb{N}[q_1, \ldots, q_n]$ will be such that $x \preceq y$ if $x_i \leq y_i$ for $i \in \mathbb{N}_n$.

Then, as a result of the discussion above, the image of language $L[k] \in \mathcal{P}_A$ under $\Phi_\varphi$ can be formally captured as the $\varphi, k$ polynomial of $L[k]$, $x$ such that

$$x \preceq \sum_{j \in \mathbb{N}} Q[\varphi, A^j] = \sum_{n \in \mathbb{N}} \left( \sum_{i=1}^{m} \alpha_i q_i \right)^j.$$

The further application of this formalization, where a lower bound on the total $\varphi$-image of a language is considered, has yet to be made. But this demonstrates a non-syntactic, quantitative way of picturing the operation of morphisms. A quantitative, rather than syntactactic consideration of distance between the morphic images of languages will be appropriate to the future investigation of the language pseudo-metrics we have discussed.
Chapter 6
Conclusions regarding distances between language; what is to be done

It is necessary now to draw some conclusions regarding what has and has not been accomplished in this dissertation. At the outset, we said that the distinctions classically made between languages had not found expression in distances. We suggested that current practical applications of formal language theory point to a need for an overall strategy of topologizing the space of formal languages. The notion of randomness was put forward as a means of testing a given topology.

In section one of this chapter we assemble the results of a systematic approach to the construction of language pseudo-metrics and the analysis of the topologies they induce on a space of formal languages. In section two, we list the questions which either remain open or which we have not addressed and advance certain proposals for further investigation. In the last section we give a final summary and close.

6.1 Conclusions regarding the Cantor, Besicovitch, and entropy topologies

We have analyzed three language pseudo-metrics, one of them a metric, and each gives interesting results. We have employed the following method.

1. Identify a language norm, i.e., a map of a language space into the non-negative real numbers, with the characteristic that the empty language maps to 0 and the norm of a language is no greater than the norm of a superset of that language.

2. Define the corresponding language pseudo-metric as the norm of the symmetric set-difference of two languages.

3. Determine basic properties of these functions, such as whether they are additive on the union of sets.
4. If the pseudo-metric is strict, map languages at distance 0 from each other to equivalence classes. Investigate the properties of these classes and the character of convergence in the resulting metric quotient space.

5. If the quotient space is not compact, try to establish an upper quotient space by identifying equivalence classes containing languages with the same norm.

6. Investigate the location of the member-families of the Chomsky hierarchy and other known characteristics of languages in the three-fold topology induced.

7. Determine the definition of randomness in the topologized language space.

In pursuing this approach, we wanted to determine whether such a method could give clear topological and linguistic results, whether different forms of randomness emerged, and what effect the use of different topologies would have on language morphisms.

The first metric was based on a finding the shortest difference between languages, the second was based on the cumulative differences between languages, and the third was based on the exponential growth rate of differences between languages. Lemmas 4.10 and 3.17, and Corollaries 3.18.1, 2.2.1, and 4.8.5 show the following.

**Lemma 6.1** The topologies $(\mathcal{P}_A, \tau_1)$, $(\mathcal{P}_A, \tau_\zeta)$, and $(\mathcal{P}_A, \tau_h)$, the Cantor, Besicovitch, and entropy topologies are mutually inequivalent, $(\mathcal{P}_A, \tau_h)$ is strictly finer than $(\mathcal{P}_A, \tau_\zeta)$, and the Cantor topology cannot be compared with the other two.

We can definitely contrast the capacity of these topologies to express language characteristics. The following examples illustrate their differences and similarities.

**Example 15** Let $A = \{a, b, c\}$. Consider the following languages over $A$:

- $L_1 = \{w \in \mathcal{P}_A : c \notin F(w)\} \in \text{LOC}_A$.
- $L_2 = \{a^n b^m : n, m \in \mathbb{N}\} \in \text{REG}_A$,
- $L_3 = \{a^n b^n : n \in \mathbb{N}\} \in \text{CF}_A$, and
- $L_4 = \{a^n b^n c^n : n \in \mathbb{N}\} \in \text{CS}_A$. 

125
These are all elements of the zero point $0_{\zeta}$ under the Besicovitch topology. Under the entropy topology, $\|L_1\|_h = \log_3 2$ and the others are in $0_h \in Q_h$. Meanwhile, the Cantor norm of each is 1, since $\lambda$ belongs to each. These results are summarized in the following tables.

\[
\begin{array}{cccc}
 d_1 & L_1 & L_2 & L_3 \\
 L_1 & 0 & 1/4 & 1/2 \\
 L_2 & 1/4 & 0 & 1/2 \\
 L_3 & 1/2 & 1/2 & 0 \\
 L_4 & 1/2 & 1/2 & 1/4 \\
\end{array}
\begin{array}{cccc}
 d_\zeta & L_1 & L_2 & L_3 \\
 L_1 & 0 & 0 & 0 \\
 L_2 & 0 & 0 & 0 \\
 L_3 & 0 & 0 & 0 \\
 L_4 & 0 & 0 & 0 \\
\end{array}
\]

\[
\begin{array}{cccc}
 d_h & L_1 & L_2 & L_3 \\
 L_1 & 0 & 0 & \log_3 2 \\
 L_2 & 0 & 0 & 0 \\
 L_3 & \log_3 2 & 0 & 0 \\
 L_4 & 0 & 0 & 0 \\
\end{array}
\]

Table 1: Language distances between small languages

Distances between languages $L_1$, $L_2$, $L_3$, and $L_4$ under the Cantor, Besicovitch and entropic pseudo-metrics.

**Example 16** Consider the following languages $L_5$, $L_6$, and $L_7$, derived from $L_2$, $L_3$, and $L_4$ of Example 15, respectively. They fall into the language families REG\_LOC, CS\_REG and CF\_CS, respectively.

- $L_5 = \{a^n b^m c A^* : n, m \in \mathbb{N}\}$
- $L_6 = \{a^n b^m c A^* : n \in \mathbb{N}\}$
- $L_7 = \{a^n b^m e^n a A^* : n \in \mathbb{N}\}$

The following table gives the results.
Table 2: Language distances between large languages

Distances between languages $L_5$, $L_6$, and $L_7$ under the Cantor, Besicovitch and entropic pseudo-metrics; also the language norms for each.

Note that the entropic pseudo-metric identified the locally testable language $L_1$, but accomplished little else. The Besicovitch distance returned very distinct values on infinite languages. The differences in the topologies does correspond to a certain expressiveness regarding language structure, but there is clearly much to be done yet in the way of a full classification of language pseudo-metrics.

6.2 Tasks yet to be finished

Foremost in the lines of research we were unable to pursue further is the possibility of new and unique spaces induced by permutative versions of each of the above-mentioned language pseudo-metrics. Distances that are invariant under permutation of symbols are free of the defect most commonly mentioned in connection with symmetric set-difference, namely that mere symbol permutation can give a near maximal distance between languages which are structurally identical. Although we showed in Corollary 1.3.2 that permutative distances exist and are language-pseudo-metrics, we discussed little of the topologies they induce.

There were two propositions integral to the investigation of language pseudo-metrics which we left as unproven conjectures.

1. The family of regular languages in dense in the Besicovitch language topological space.
It was, however, shown, in Lemma 3.19, that regular languages are dense in the upper Besicovitch quotient space, and, in Lemma 4.10, that regular languages are dense in the upper entropic quotient space.

2. Every language code (injective language morphism generated by a literal mapping) is continuous on its image.

It was shown, in Lemma 5.2, that a language code is continuous on its image if it is a map from a topologized space to a space with a finer topology.

Additionally, the investigation into randomness on the entropic topology was transferred to randomness on the upper quotient space, with the unsatisfactory result that all languages sufficiently large might be regarded as nonrandom. An adjustment to this approach is clearly in order.

6.2.1 Distinguishing regular languages metrically

The distinction of certain subfamilies of the regular languages by metrics has been demonstrated in that they have been forced into the same equivalence class and therefore cannot be dense in the topologized language space. But it remains to accomplish a similar distinction by metric means between languages in $\text{LIN}_A$, for instance, and languages in $\text{REG}_A$. A sketch of one possible attack on this problem is presented here.

The idea is to construct a metric designed with syntactic goals. To make certain that regular languages are differentiated by this metric from non-regular languages, take advantage of the Myhill-Nerode theorem in the following manner. Let $d_R(L, M)$ be given by ration of the number of follower sets in $L \triangle M$ to $k$, with the limit supremum taken as $k$ goes to infinity.

**Definition 6.2.1** Let $R_k(L) = \# \{ w \in A^* : uw \in L \} : u \in L \}$ and let

$$\|L\|_R = \lim_{k \to \infty} \sup \frac{R_k(L)}{k}. \quad (6.1)$$

Then, as usual, let

$$d_R(L, M) = \|L \triangle M\|_R.$$
This is now a generalized metric, with $||L||_R = \infty$ for certain non-regular languages $L$, and $||L||_R = 0$ if $L$ is regular. Since this steps beyond the bounds set in this paper for pseudo-metrics, it is an open question whether $d_R$ satisfies other intuitive requirements for usefulness as a language distance.

This sort of “brute-force” approach may be improved upon by considering a different aspect of languages, as posed in the following subsection. Similarly, the implementation of topological entropy weighted by the word-length at which a factor first appears in a language could lead to a separation of non-RE from RE languages.

6.2.2 Distances between grammars

The definition of languages via syntax grammars already implies that a quotient space similar to those used in this paper is in operation. A minimal number of rules of production must be applied to go from the axiom to any word in the language, and the iteration count implies equivalence classes of subsets of the language. To develop a language distance, it may not be necessary to look any farther than the grammars themselves. There is, for instance, a fundamentally quantitative distinction between CS and non-CS languages: every element of the production relation is a map of a word to a word that is no shorter (with the exception of the axiom).

This implies that a metric based on the ratio of word lengths in the grammar can separate CS and non-CS languages. It needs to be established that this gives a pseudo-metric on languages, however, in the sense defined in this paper, or in some other sense agreeable to the intuitive understanding of languages. The full power of giving a polynomial representation to a morphic image, as described in Chapter Five, would then be instrumental in defining the topology induced by a syntax metric.

6.2.3 Algebraic topology on the space of languages

It was also mentioned at the outset that algebraic topology has found an application in areas of language and computation theory. Even in DNA computation and the analysis of DNA function, homology and knot-theory are eminently useful. This dissertation has not addressed the issues of the connectedness and possible path-connectedness of the Besicovitch and entropic topologies; of course, the Cantor topology is totally disconnected.
The possibility exists of defining simplicial complexes of languages, the faces of which should correspond to classes of words, where the language space is connected. This may be a more complete way of capturing language structure in a metric.

6.2.4 The $d_L$ metric

As mentioned in Chapter 2, there are significant variations on the Cantor metric. One of them, developed for use with languages over sequences, can be adapted to a formal language space. In doing so, we refer primarily to the work of Calude, et al. in [5]. The adaptation, which needs completion and further topological analysis, we present without proof.

Given $F \subseteq \text{FIN}_A$, let $F^\delta$ be the set of languages $L \in \mathcal{P}_A$ such that there is a subset $F_L = \{F_i\}_{i \in \mathbb{N}}$ of $F$ such that $F_i \uparrow L$. Note that $\text{FIN}_A^\delta = \mathcal{P}_A$, and that if $F$ is finite, $\text{FIN}_A^\delta = F$.

Let $F \subseteq \text{FIN}_A$ be a set of finite languages. Define the distance $d_F : \mathcal{P}_A \times \mathcal{P}_A \to \mathbb{R}_{\geq 0}$,

$$d_F (L, M) = \begin{cases} 0, & \text{if } L = M, \text{ and} \\ 2^{-\min \{ k : L \Delta^k M \neq \emptyset \text{ or } L \Delta^k L' \neq \emptyset, \text{ for some } L' \in F^\delta \}}, & \text{otherwise.} \end{cases} \quad (6.2)$$

Note that $d_{\text{FIN}_A} = d_1$. Open balls of radius $\varepsilon$ around a language $L \in \mathcal{P}_A$, given $L \subseteq \text{FIN}_A$, consist of the following:

$$B_{\varepsilon, F} (L) = \begin{cases} \{L\} & \text{if } d_F (L, M) \geq \varepsilon, \forall M \in \mathcal{P}_A \setminus L \\ \mathcal{P}_A & \text{if } \varepsilon \geq 1, \text{ and} \\ \mathcal{C}_{F, L, \varepsilon} & \text{otherwise.} \end{cases} \quad (6.3)$$

where $F_L = \{F \in F : \# (F \triangle^{|\log_2 \varepsilon|} M) = 0\}$.

The most appropriate designation for this topology is $\tau_{F^\delta}$. This is due to the conclusion from [4] that, over languages of infinite sequences, a similarly-defined $F^\delta$ topology is equivalent to some $G^\delta$ topology if and only if $F^\delta = G^\delta$. For otherwise, without loss of generality, there is a language $L \in \mathcal{P}_A$ which is the $d_1$ limit of a subsequence of $F$, but of no subsequence of $G$. Then there exists $N \in \mathbb{N}$ such that $m > N$ implies $L \triangle^k G \neq \emptyset$ for all $G \in G$. But an open neighborhood of radius $\varepsilon$ around language $L$ would then be the singleton $\{L\}$, in the $F^\delta$ topology, whereas an open
neighborhood of radius \( \varepsilon \) around \( L \) in the \( G^{\delta} \) topology would be the cylinder set \( C_{L,\lceil \log_2 \varepsilon \rceil} \), as in \( \tau_1 \).

Many of the conclusions in [4] apply without difficulty to \((P_A, \tau_L)\), for \( L \subseteq \text{FIN}_A \). We present them here without proof, for further investigation.

1. The metric space \((P_A, d_L)\) is not necessarily compact.

2. Let \( \overline{L} \) denote the closure of \( L \subseteq P_A \) in \((P_A, d_1)\). Then the closure in \((P_A, d_L)\) of \( L \subseteq P_A \) is the set \( \overline{L}^{\delta} = \overline{L} \cap (L \cup F^{\delta}) \) (Corollary 6, [4]).

3. Let \( \mathcal{I}_F \) denote the set of all isolated points in \((P_A, d_F)\). Then the following are equivalent:

   (a) The set \( \mathcal{I}_F \) is dense in \((P_A, d_1)\);

   (b) The set \( F^{\delta} \) is nowhere dense in \((P_A, d_F)\); and

   (c) The set \( F^{\delta} \) is a maximal nowhere dense set.

6.3 Conclusion

The space of languages proves to be rich in material and relevant questions from the standpoint of topology, morphisms, and randomness. Language collections are quantifiable as objects and in relation to each other. The approach of giving a rigorous definition of a language pseudo-metric, by supplementing ordinary pseudo-metric conditions with the requirement that languages with a greater number of distinct words be farther apart proves to make the structures within language collections more apparent. By Proposition 1, each language pseudo-metric may be associated with language norm, i.e., a definition of the relative size of a language. This has shown interesting results and raised intriguing questions. Under a Cantor distance, language spaces are totally disconnected; they are all copies of the Cantor space. By Theorem 2.1 the non-recursively enumerable languages are the random languages in the Cantor space. Under a Besicovitch distance the language space is connected, non-compact, and can be partitioned into equivalence classes with unique antipodes. We used Proposition 2 to show that, under a second quotient map, the unit interval appears as the compact quotient of the space. We adapted topological entropy into an entropic distance, and showed, in Corollary 4.8.5, that entropic distance induces a topology which is a refinement of the Besicovitch topology. We found a subfamily of the regular languages which is dense in this space,
but not in the Besicovitch topology. This is possible because the languages with entropic norms less than one present an enlarged picture of a portion of the zero point in the Besicovitch quotient topology. It is not difficult to propose additional topologies which separate the finite languages from the rest of the language space, but it is unknown whether a meaningful metric can be achieved that separates the various language families of the Chomsky hierarchy from each other.

The language norms corresponding to these pseudo-metric act to filter out specific aspects of languages. The big picture of where languages lie is given by the Besicovitch norm, which, more or less, determines proximity to ideals of $A^*$; all locally testable languages belong to the same equivalence class. The entropic norm refines this picture near the zero point to give indication of a locally testable languages. If the language is finite, the Cantor norm or a variation of it can give a finer evaluation of the relative size of the language. This ordering of partially comparable and incomparable topologies by the linguistic conclusions of which they are capable suggests the possibility of further positive results from the systematization of the space of language pseudo-metrics.

Determining the precise notion of randomness defined by a distinct language topology has proven to not always be straightforward. It is convenient, we have seen, when upper quotient spaces are found to be homeomorphic to the unit interval. Improved methods are suggested in the case of the entropic topology. The volumes of literature on word morphisms exist (see [14]), contrasts sharply with the dearth of investigation of language morphisms. We have shown some interrelationships between classes of morphisms, shown special features of classes of metrics under certain language pseudo-metrics, advanced a matrix-form representation of the pre-image of a morphism, and a formal polynomial representation of the image of a morphism. The tentative steps seem to set the stage for more substantial research.

The recent developments in various scientific disciplines which have led to an unprecedented increase of the ability of researchers to assemble stores of data, using labels appropriate to the field to identify physical events, is a fundamental achievement of the alliance of computer technology with empirical investigation. Catalogs of symptoms, genomic sequence libraries, human intercommunication limitlessly expanding, real-time pressure-velocity data from the eye-wall of a hurricane — all of these present themselves for our inspection as non-commutative output of one or another discrete time dynamical system. Some of these are collections of words of unimaginably length, languages
in which the information of Nature is expressed, most with naturally associated notions of proximity and size. Wherever such data are found, they not only express a condition, but also imply a potential and, in this sense, encode the pre-conditions for the following event. To attempting to forecasting events from within the symbolic dynamics of the system is therefore to inquire into the topology of languages; it is at heart a mathematical task (for additional examples of this, see [1, 21, 31, 32]).

This paper has been an experiment in the systematic assessment of language topology, morphisms, and randomness. Our approach, based upon symbolic dynamics and certain past and recent work in language theory ([5, 38]) has been motivated by the conviction that progress in this direction lays the groundwork for the comprehension, through modelling by languages, of phenomena in many fields. The Besicovitch and entropic topologies are to be regarded as first entrants in a series of structures enabling deeper understanding of language dynamics and an investigation which produces tools, along the way, such as the geometrically complex quotient spaces involved, for the more general attack on the space of language pseudo-metrics.
References


guages.” Aspects of Molecular Computing, Essays Dedicated to Tom Head on the Occasion of


Enforcing Systems.” Preprint.


[16] P. Hertling and K. Weihrauch. “Random elements in effective topological spaces with mea-

the 7th International Meeting on DNA Based Computers. N. Jonoska and N. C. Seeman, eds.


[19] N. Jonoska and K. Mahalingam. “Languages of DNA based codewords.” Preliminary Pro-
ceedings of the 9th International Meeting on DNA Based Computers. J. Chien and J. Reif, eds.
(2003) 58–68

Developments in Language Theory. C.S. Calude et. al., eds. Lecture Notes in Computer Science


[24] W. Kuich. “Semirings and Formal Power Series: Their Relevance to Formal Languages and


[27] S. MacLane. Categories for the Working Mathematician. Springer-Verlag. New York, NY,
1971.

[28] A. Mateescu and A. Salomaa. “Aspects of Classical Language Theory.” Handbook of For-
mal Languages, Vol.1: Word, Language, Grammar. G. Rozenberg, A. Salomaa, eds. Springer-

619.


[31] J. Roca, D. W. Sumners, M.E. Vázquez. “Novel display of knotted DNA molecules by two-

Notices of the AMS. 42:5 (May 1995) 528–537.


About the Author

David Kephart was born in the Pacific Northwest on the third Sunday of a wintry month in the middle of the twentieth century. His interest in mathematics was stirred by the books of his parents, Donald and Silva, and especially encouraged by his engineer-grandfather Lewis H. Austin. Despite this, he took a long hiatus from academe and only began his studies in the field in 1996. Upon completing lower-division work at San Francisco City College, he moved to Tampa, Florida. He was awarded his Bachelors of Arts, magna cum laude, and Masters of Arts degrees in Mathematics simultaneously at the University of South Florida in August, 2001.

Since then, he has been engaged in research under the direction of Nataša Jonoska in fields of DNA codeword generation and the general questions in language theory and symbolic dynamics related to this. He is the co-author of two published articles on the subject and of a program, C\textsc{ODE}G\textsc{EN}, which accomplishes codeword generation. His interests lie in the use of formal languages and symbolic dynamics to model the behavior of real-world phenomena.