Universiteit
Leiden
The Netherlands

# Affordances and limitations of algorithmic criticism

Verhaar, P.A.F.

**Citation**
Verhaar, P. A. F. (2016, September 27). *Affordances and limitations of algorithmic criticism*. Retrieved from https://hdl.handle.net/1887/43241

| | |
|---|---|
| Version: | Not Applicable (or Unknown) |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/43241 |

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/43241 holds various files of this Leiden University dissertation

**Author:** Verhaar, Peter
**Title:** Affordances and limitations of algorithmic criticism
**Issue Date:** 2016-09-27

Chapter 6

# Data creation

## 6.1. Introduction

The previous chapter concentrated on the question whether or not computational methods can beneficially aid scholars in unravelling the manifold linguistic and rhetorical aspects that can give literary works their unique quality. The aspiration to automate some of the core activities within the context of literary criticism generally stems from the expectation that digitisation can yield particular advantages with respect to speed, precision, or quality. This impulse to digitise is based additionally on the conviction that particular core activities are indeed amenable to digitisation. In his widely cited essay "Computing Machinery and Intelligence", Alan Turing has stressed that the value of the computer, as a universal machine, lies in the fact that it can "mimic any discrete-state machine",[378] provided that the functions of this machine can be described via a finite number of unequivocal activities. When scholars take "the computational turn",[379] they essentially need to consider if sections of their existing scholarly practices can be formalised through algorithms. Research projects in the field of literary informatics aim to address the questions which are asked traditionally within literary studies in innovative ways by making use of computational methods. To ensure that computation can veritably be supportive of the discipline in which the tools are adopted, it is necessary to take the existing practices and traditional interests as the point of departure, and to manipulate the toolset in such a way that it can be used, as much as is possible, in the service of these traditional objectives. Matching the possibilities that are offered by informatics to existing scholarly objectives often demands many efforts and much creativity, however. The digital medium principally provides support for calculations, and it can process data only if these explicit and unequivocal. Because of these demands, it is not always possible to represent conventional scholarly approaches as quantitative operations.

The digital heuristics that emerge eventually are often an amalgamation, resulting from a negotiation between what is desirable from a scholarly perspective on the one hand and the limitations and the affordances of computation on the other. On the one hand, Turing's conceptualisation of computing implies that digital technology is malleable. The universal machine was not built for a unitary

[378] Alan Turing, "Computing Machinery and Intelligence", in: Noah Wardrip-Fruin & Nick Montfort (eds.), *The New Media Reader*, Cambridge Mass.: The MIT Press 2003, p. 54.
[379] Caroline Basset, "Canonicalism and the Computational Turn".

objective, but can be transmuted by its users to make it serve particular purposes. There may also be a backlash, however. While it may be true that we can influence the nature of the technologies we use, these technologies may simultaneously, perhaps unconsciously, shape us.

For the purpose of this thesis, the concrete obstacles in aligning traditional practices and scholarship based on data processing have been investigated in a practical manner, by conducting a case study. The case study consisted, more specifically, of a computer-based analysis of the poetry of the Northern Irish poet Louis MacNeice. Robin Skelton notes that Louis MacNeice is "one of the master craftsmen" in English poetry and also stresses his poems are "essential reading for anyone who cares to study or to practice the intricacies of English verse".[380] MacNeice's poetry simultaneously exemplifies the sundry obstacles that may emerge from the manifold interpretability of language. Reflecting on his own verse craft, MacNeice wrote that language is generally "a traffic in symbols" and added that "these symbols are plastic - an endless annoyance to the scientist but God's own gift to the poet".[381] Interestingly, as will be shown, MacNeice's main thematic concerns mirror some of the central challenges in the field of literary informatics.

As was indicated in the previous chapter, computational analyses of texts are often based on prior quantifications of low-level linguistic features, such as the most frequent words or occurrences of specific grammatical constructions. Studies which make use of conventional text tools, and which are consequently based on such formal characteristics, are, to some extent, intellectually remote from traditional forms of literary scholarship. In the case of poetry research, the focus is generally on the description and the interpretation of aspects such as meter, figures of speech, imagery or themes. As part of the case study, a basic software application was developed for the automated detection of a large number of literary devices. Unless indicated otherwise, all software has been programmed by myself.[382] Where possible and where relevant, the studies and the functionalities which have been discussed in Chapter 2 have been used as guidelines.

The case study was divided into three phases. First, an analysis was made of a large part of the existing criticism on MacNeice. An inventory was compiled of some of the questions that were asked about the verse, and of the methods that were used to answer these questions. Most pertinently with respect to the case study's central aims, data was also collected about the various literary devices that have been identified by critics. A second phase focused on the practical obstacles involved in the creation of data about literary devices. The observations which were

---

[380] Robin Skelton, "Celt and Classicist: The Versecraft of Louis MacNeice", in: Terence Brown & Alec Reid (eds.), *Time Was Away*, Dublin: Dolmen Press 1974, p. 53.

[381] Louis MacNeice, "Experiences with Images", in Alan Heuser (ed.), *Selected Literary Criticism of Louis MacNeice,* Oxford: Clarendon Press 1987, p. 155.

[382] All the code that was developed for the purpose of this dissertation can be found at <https://github.com/peterverhaar/Phd>

made in the existing criticism formed inspiration for the design of the text analysis software. The case study aimed to investigate whether or not the statements made by critics could also be produced on the basis of text mining tools. In a third phase, the data that were produced were also analysed statistically. One of the objectives in the third phase was to explore if data processing may help to produce new readings and if these can contribute to an improved understanding of the texts. The aim was to design analytic procedures which can help to address the issues which were also raised in the existing criticism. The case study is largely based on the hypothesis that specific aspects can be investigated more thoroughly and more consistently if they are analysed through computation. The case study also investigated whether or not computers can enable scholars to answer existing questions differently, or, more ambitiously, if digital methods can also allow critics to ask entirely new types of questions.

The following section firstly gives a brief summary of the existing criticism. For the largest part, however, section 6.1. discusses the difficulties connected to the creation of data. This chapter discusses the various functionalities that have been programmed in close conjunction with the reasoning that has been following during the implementation of these functionalities. Chapter 7 explains how the data that have been created are stored, focusing, more particularly, on the benefits and the disadvantages of a number of data formats. Chapter 8, finally, focuses on the broader patterns that can be generated via analyses of the resultant data.

## 6.2. Case Study

### 6.2.1. Introduction

To clarify the general background of the central case study, this section offers a brief discussion of the life and work of Louis MacNeice, together with a summary of the main issues that are addressed in the various critical studies of MacNeice's work. MacNeice was born in 1907 in Belfast, and grew up in Carrickfergus, where his father was a Church of Ireland rector. MacNeice's mother died when he was five years old. Since the age of ten, he was educated at public schools in England and at Merton College in Oxford. In 1929, MacNeice started working as a lecturer in Classical literature at Birmingham University, and, during this period, he also published his first volumes of poetry. MacNeice became a highly prolific author, writing ten volumes of poetry, a large number of radio plays and a sizable body of literary criticism. During his lifetime, MacNeice was mainly known as a member of the group of thirties poets, which also included W.H. Auden, Cecil Day Lewis and Steven Spender.

After MacNeice's early death in 1963, a large number of critical assessments have been made of his full oeuvre and if its impact.[383] Most critics agree that MacNeice's literary output can broadly be divided into three phases. The poet initially rose to fame in the early 1930s following the publication of his first volumes *Poems* and *The Earth Compels*. The poems in these volumes have frequently been lauded for their strong sensual and visual qualities, the skilful use of meter and rhyming schemes, and the overall colloquial and free-flowing register that is adopted. The long poem *Autumn Journal*, largely written in 1938, is frequently considered MacNeice's masterpiece. The early 1950s, however, during which MacNeice wrote *Ten Burnt Offerings* and *Autumn Sequel*, are generally viewed as a period of creative impasse. About *Ten Burnt Offerings*, Allan Gillis writes that "the words fail to animate, and the verse remains stubbornly flat-lined".[384] In the second half of the 1950s, MacNeice gradually began to develop a new style, in which he experimented with parable and with different types of rhymes. A number of scholars have noted that the last three volumes *Solstices*, *Visitations* and *The Burning Perch* contain some of MacNeice's best poetry. Goodby stresses that, in this third and final phase, the poet managed to achieve a "lyric compression, revealing an adjustment to the darker climates of Cold War and middle age".[385]

Terence Brown argues that MacNeice's verse is informed fundamentally by a profound scepticism, and by a refusal to commit fully to any particular philosophy or religious creed. Brown connects this scepticism to biographical factors, such as the early loss of his mother, and to his state of living as an exiled Irishman in England.[386] The first volumes in particular represent the world of sensual experiences as intrinsically complex and transient, and stress that it is impossible to explain or to understand such plurality through a finite set of values or rules. Derek Mahon stresses similarly that MacNeice's verse reflects a clear distrust of generalisations. Throughout his literary career, however, the poet was beset simultaneously by a fear that a lack of commitment and the consequent superficiality cannot be sufficient, and he continued to hanker after more permanent and more profound moral values. The poetry, in short, displays a clear tension between profundity and superficiality.

---

[383] Important book-length studies were written by Terence Brown, John Stallworthy and Edna Longley. Brown and Longley have also edited collections of essays, such as *Studies into Louis MacNeice* and *Time was Away*.

[384] Alan Gillis, ""Any Dark Saying": Louis MacNeice in the Nineteen Fifties", in: *Irish University Review*, 42:1 (20 May 2012), p. 108.

[385] John Goodby, "Louis MacNeice", in: David Scott Kastan (ed.), *The Oxford Encyclopedia of British Literature*, Oxford: Oxford University Press, p. 365.

[386] Terence Brown, *Louis MacNeice: Sceptical Vision* (Dublin; New York: Gill and Macmillan 1975), p. 20, and passim.

Various scholars have also noted that MacNeice's basic scepticism and agnosticism had affected his political orientations. While contemporaries such as Auden, Spender and Day-Lewis clearly reacted to the social and the political climate of the 1930s by writing verse that was overtly socialist, Peter McDonald notes, by contrast, that MacNeice's 1935 volume *Poems* seems "aggressively neutral".[387] MacNeice most frequently addresses politics and social commitment in an indirect manner. His main interest was in the issue of "how far the self is able to marginalize the other into mere 'context' and how far it is the context, the other, which gives meaning to the self". MacNeice's leanings, nevertheless, were predominantly left-wing.[388] After his travels to Iceland and to Spain at the onset of the civil war, MacNeice clearly became more politically engaged.[389] Gillis observes that *Autumn Journal* "fused the personal and the communal in a poetics of social awareness and commitment".[390] The disillusionment that followed the Second World War, however, appears to have resulted in a disinterest in political involvement and in misgivings about the value of social commitment. In MacNeice's later work, "the sense of both self and society has become morbidly phantasmagorical".[391]

MacNeice's troubled relation with Ireland has also attracted much critical interest. The poet had left his native country at an early age, and, throughout his lifetime, Ireland remained a place both to admire and to reject. While many poems idealise the West of Ireland, texts such as "Valediction", "Neutrality" and "Autumn Journal" also contain vehement reactions against Irish politics and Irish culture. Importantly, however, MacNeice has had a profound impact on the poetry written in Ulster in the 1970s and the 1980s. Heather Clarke records that, while MacNeice was viewed largely as a minor poet working in the shadow of W.H. Auden during his lifetime, the poet's reputation was restored posthumously by Northern Irish poets such as Seamus Heaney, Derek Mahon and Michael Longley, who viewed MacNeice as a major figure in the poetic tradition of Ulster and who recognised him as a "model of cultural transience and displacement".[392] Clarke also observed, however, that this repatriation was also a reinvention, as these poets needed to ignore MacNeice's political commitment during the Second World War. Derek Mahon writes that, while the poetry is often not specifically about Ireland, MacNeice nevertheless has "some sort of Irish sensibility".[393] John Goodby argues

[387] Peter McDonald, "The Falling Castle", in: Jacqueline Genet & Wynne Hellegouarc'h (eds.), *Studies on Louis MacNeice*, Caen: Presses universitaires de Caen 1988, s. 1.

[388] Goodby notes that MacNeice largely adopted "a pragmatic, skeptical socialism". See John Goodby, "Louis MacNeice".

[389] It is clear from his discussion of the role of poetry in society in the essay *Modern Poetry*, and it is also apparent in a large number of poems in *The Earth Compels*, and, most notably, in *Autumn Journal*.

[390] Alan Gillis, ""Any Dark Saying": Louis MacNeice in the Nineteen Fifties", p. 107.

[391] Ibid., p. 106.

[392] H. Clark, "Revising MacNeice", in: *The Cambridge Quarterly*, 31:1 (2002).

[393] Derek Mahon, "MacNeice in England and Ireland", in: Terence Brown & Alec Reid (eds.), *Time Was Away: The World of Louis MacNeice*, Dublin: Dolmen Press 1974, p. 113.

similarly that MacNeice addressed quintessentially Irish themes, and that his "interest in relativity and flux, his constant attempts to deconstruct binaries, his concern with the self and with tradition as potential self-betrayal make him an exemplary Irish writer".[394]

This study is interested in the algorithmic investigation of literary devices, and, for this reason, it was particularly important to collect critical descriptions of the way in which MacNeice expressed these themes through language. Many critics have commented on the relaxed nature of the meter, the elegance of the rhyming schemes, and the frequent use of devices such as alliteration and assonance. About the use of rhyme, MacNeice argues, in his long essay *Modern Poetry*, that perfect rhymes can add musicality to the verse, while it also suggests an insincerity. As a compromise, poets can deploy variations, such as "internal rhyme, off-rhymes, bad rhymes and 'para-rhymes'" or they can rhyme "a stressed against an unstressed syllable".[395] In his own work, MacNeice experimented with the many similarities in sounds and with various types of rhyme. Skelton stresses that the poet frequently deployed "elaborate patterns of near-rhyme, assonance and consonance",[396] and Longley has similarly commented on the "importance of emphatic rhythmical punctuation — like assonance, internal rhyme and refrain".[397] In a close reading of the poem "Donegal Tryptych", Terence Brown notes that "the use of assonance and alliteration, the sheer vigour of the diction, together with the patterning, all draw attention to the language as an almost physical object".[398]

Various critics have focused on the use of repetitions. Neil Corcoran notes that the poem *Leaving Barra* contains various forms of repetition, such as a reiteration of the exact same word in two consecutive lines, the use of an identical phrase at the beginning and at the end of the poem ("the dazzle on the sea") and a repetition of parts of words in other words on the same line ("the rain and the rainbow"). Next to being aurally pleasing, such repetitions are also "thematically functional",[399] as such repetitive elements are "sensitively mimetic of the mind in progress – self-scrutinising, self-corrective, advancing hesitantly but keeping moving".[400] Allen Gillis observes similarly that MacNeice's poetry frequently contains "repetitive riffs" and that the poet often uses "chiasmus and chiastic-like effects".[401] Interestingly, such chiastic repetitions create a "paradox of movement and stasis", and

[394] John Goodby, "Louis MacNeice".
[395] Louis MacNeice, *Modern Poetry: A Personal Essay*, p. 131.
[396] Robin Skelton, "Celt and Classicist: The Versecraft of Louis MacNeice", p. 43.
[397] Edna Longley, "Louis MacNeice: Aspects of His Aesthetic Theory and Practice", in: Jacqueline Genet & Wynne Hellegouarc'h (eds.), *Studies on Louis MacNeice*, Caen: Presses universitaires de Caen 1988, s. 1.
[398] Terence Brown, *Louis MacNeice: Sceptical Vision*, p. 172.
[399] Neil Corcoran, "The Same Again? Repetition and Refrain in Louis MacNeice", in: *The Cambridge Quarterly*, 38:3 (4 August 2009), p. 216.
[400] Ibid., p. 214.
[401] Alan Gillis, ""Any Dark Saying": Louis MacNeice in the Nineteen Fifties", p. 111.

they convey the way in which things "are both singular and multiple at the same time".[402]

Robin Skelton, in his essay "Celt and Classicist: the Versecraft of Louis MacNeice",[403] argues that MacNeice's Irish background clearly transpired via the usage of poetic devices which are characteristic of Celtic verse. MacNeice often created lines with an uneven number of syllables and alliteration, near-rhyme, assonance and consonance. In a close reading of the poem "Aubade", Skelton demonstrates that a Celtic influence is noticeable in the use of deibhide rhyme, in which unrhymed syllables rhyme with stressed syllables.[404] There is also a preference for slant rhyme over perfect rhyme, as the poem stresses the acoustic similarities between the words "apple" and "happy" and between "dawn" and "war". Skelton also discusses various occurrences of internal or Aicill rhyme. A specific case of Aicill rhyme occurs when the consonants of the final word of one line are repeated in the consonants of the first words on a line that follows. This device is used in "Order to View", in which the consonants in the word "crypt" on line 7 are repeated in the word "empty" on line 9.[405]

The body of literary criticism is extensive, and it is impracticable to include a complete discussion of the debate on the merits and the shortcomings of MacNeice's writings in this section. This section principally aimed to highlight a number of topics, and to identify a variety of questions that can be addressed further in the case study. Critics of MacNeice's poetry have focused, in short, on the stylistic and thematic differences between the separate volumes, the impact of MacNeice's Celtic background, MacNeice's own literary influences and the poet's influence on later authors, and the nature and the function of refrains and other forms of repetition. In analyses of MacNeice's language, critics have emphasised the use of slant rhyme, alliteration, assonance and Celtic devices such as deibhide and aicill rhyme. Corcoran and Gillis have also commented on the poet's systematic repetition of words, parts of words or of specific groups of words.

### 6.2.2. Basic annotations

The text corpus was created by scanning the eleven volumes of poetry that were selected. The case study focuses on *Poems*, *The Earth Compels*, *Autumn Journal*, *Plant and Phantom*, *Solstices*, *Holes in the Sky*, *Autumn Sequel*, *Ten Burnt Offerings Visitations*, *Springboard* and *The Burning Perch*. Machine-readable versions of the poems were obtained through Optical Characters Recognition (OCR). The results of the OCR were proofread to ensure that the electronic texts were free of scanning errors. In total, the corpus consisted of 311 poems, which

[402] Alan Gillis, ""Any Dark Saying": Louis MacNeice in the Nineteen Fifties", p. 113.
[403] Robin Skelton, "Celt and Classicist: The Versecraft of Louis MacNeice".
[404] Ibid., p. 43.
[405] Ibid., pp. 44-45.

collectively contain 144269 tokens and 16623 types. This data set is not extensive, if measured against the criteria mentioned in Doug Laney's article on big data collections.[406] The relatively modest size of the corpus has had the advantage that the results produced by text analysis tools could be evaluated effectively by inspecting the original texts. A basic form of TEI encoding was also added to the plain texts. A program was developed in the PERL programming language to encode the texts strings that were delineated by hard returns in the original text file as verse lines, using the <l> element. All lines were also numbered. Each line received an identifier, which was captured in the @n attribute of the <l> element. POS tags and lemmas were added using the Morphadorner application, which was developed as part of the MONK project.[407]

A method was also developed for the creation of phonetic transcriptions of all the verse lines. The PoetryAnalyzer tool, which was developed by David Kaplan, could not be used, as it was based on a dictionary of American pronunciation.[408] A similar tool was developed by making use of the pronunciation dictionary that was developed for the MRC Psycholinguistic Database, which is available in its entirety from the Oxford Text Archive.[409] The format of the transcriptions from this dictionary was converted to SAMPA, which is a phonetic script based on the International Phonetic Alphabet, which makes use of ASCII characters only. Since it was found that the types that occurred in the corpus were not all available in the selected dictionary, a number of algorithms have also been implemented for the creation of phonetic transcriptions for the remaining tokens.[410] These algorithms, unfortunately, did not function perfectly, mainly because of the fact that there is no strong connection between spelling and pronunciation in the English language.[411]

---

[406] As one of the criteria, Laney states that data collections can be considered big if they are too voluminous to be managed on a single computer of average capacities. This was clearly not the case for the data produced in this study, as the total size does not exceed 50 MB. See Douglas Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, (2001). As the data collections that humanities scholars work with generally have a modest volume and a low velocity, it may be argued that the term "big data", as conceptualised within computer science and data science, is not fully applicable within the humanities. As an alternative to the term, Allen Riddell has coined the phrase "very large collection" to denote a collection which "contains more texts than a single researcher would be expected to digest in a year's worth of dedicated reading". See Allen Riddell, "How to Read 22,198 Journal Articles: Studing the History of German Studies with Topic Models", in: Matt Erlin (ed.), *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, Boydell & Brewer Ltd 2014, p. 92.

[407] <http://morphadorner.northwestern.edu> (16 February 2014)

[408] David Maxwell Kaplan, *Computational Analysis and Visualized Comparison of Style in American Poetry*.

[409] M.D. Wilson, "The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2", in: *Behavioural Research Methods, Instruments and Computers*, 20:1 (1988). The data set can be obtained at <http://ota.ox.ac.uk/desc/1054> (10 February 2013).

[410] The algorithms were based on the pronunciation guidelines that were provided in Beverley Collins & Inger Mees, *The Phonetics of English and Dutch* (Leiden: Brill 1998).

[411] Edward Carney, *A Survey of English Spelling* (Routledge 1994).

Mistakes in the transcriptions that were generated algorithmically were corrected manually.

The fragment below given an impression of how the basic linguistic annotations have been captured in TEI. The example is first line of the poem "MeetingPoint".

```
<l n="1">
<w phon="!taIm" lemma="time" pos="NNP">Time</w>
<w phon="wQz" lemma="be" pos="VBD">was</w>
<w phon="@-!weI" lemma="away" pos="RB">away</w>
<w phon="@n" lemma="and" pos="CC">and</w>
<w phon="!sVm-we@" lemma="somewhere" pos="RB">somewhere</w>
<w phon="!els" lemma="else" pos="RB">else</w>
</l>
```

The data that were produced in this way provided the basis for additional forms of processing, which focussed on the detection of specific literary devices. The methods that were developed are discussed in more detail in the following section.

### 6.2.3. Perfect rhyme, slant rhyme and semi-rhyme

Many of MacNeice's poems contain highly sophisticated experiments with poetic conventions and with the artistic effects that can be achieved through skilful combinations of similarities in sounds. Via computation, a number of the aural structures can be explored in a structural manner. Firstly, I have developed an algorithm for the identification of occurrences of perfect end rhyme. David Kaplan's PoetryAnalyzer software detects cases of perfect rhyme by firstly extracting the final stressed syllable of the last word of each line, and by removing all consonants at the start of this syllable. In the case of feminine rhyme, this final stressed syllable is also followed by additional unstressed syllables. If specific phoneme sequences created in this way occur more than once, this is considered to be an instance of perfect rhyme. A similar logic was implemented in the application for this thesis' case study. A number of modifications were found necessary, however. In Kaplan's software, the algorithm operates on a window of four lines. Since it was found that MacNeice often uses rhyming schemes that span more than four lines, it was decided to extend the size of the window and to examine repetitions of rhyming schemes in full length stanzas. As rhyming schemes are generally defining at the level of stanzas, rhymes across different stanzas were disregarded. MacNeice has also written various poems which essentially consist of one lengthy stanza, such as "Valediction". Such long poems have been divided, somewhat arbitrarily, into clusters of 8 lines. I assumed that rhymes in a larger number of lines were not close enough to be heard together. Using the procedure that was discussed, perfect rhyme was found in the words "implications" "revelations" in "A Contact", "scales"

and “tails” in “Sunday Morning”, “wells” and “cells”, “cease” and “caprice” in “April Manifesto”, and “halt” and “salt” in “Carrickfergus”. The procedure that was implemented also returned the rhyming scheme for each set of lines.

Further tests exposed a number of additional complications. The algorithm initially overlooked a number of lines which appeared to contain obvious examples of perfect rhyme. The poem “The Return”, for instance, contains lines ending in the nouns “consummations” and “patience”. The software application transcribed the former word as /*kɒn-sə-ˈmaɪ-ʃənz/ and the latter word as /ˈpəɪ-ʃəns/. Because of the different final phonemes, these line endings were considered to be sonically distinct. The final phoneme sequences sound very similar, however, and it seems incorrect to label this particular example as an instance of slant rhyme. It was decided to implement a somewhat less rigid form of matching, and to represent all sibilant sounds as a single /s/. Similarly, the diphthongs /əʊ/ and /ɪə/ were replaced, respectively, with /ɔ:/ and /i:/. Using this more lenient form of matching, a number of additional instances of perfect rhyme were found. Examples include the lines ending in “sprawls” and “Gauls” in “Museums”, “listen” and “vision” in “Visitations” and “choose” and “booze” in “Alcohol”.

Examinations of the results also revealed that MacNeice frequently uses repetitions of the exact same words at the end of poetic lines. It was decided that, if different lines contain identical words, this does not constitute perfect rhyme. In a number of cases, such repetitions may be considered instances of rime riche, as they consist of homonyms in which the spelling and sound are fully alike, while there is a significant difference in the meaning of these words.[412] In “Eclogue by a Five-Barred Gate”, for example, the poet writes that “My dream will word well / But not wear well / No dreams wear at all as dreams / Water appears tower only while in well” (ll. 85-87).

A similar method was implemented for the recognition of half rhyme, which, as was noted by Skelton, is an important and recurrent form of rhyme in MacNeice’s verse. The *Oxford Dictionary of Literary Terms* defines half rhyme, or slant rhyme, as a form of rhyme in which there is an agreement in the final consonants of the final stressed syllables, and a difference in the vowel sounds. An alternative form of slant rhyme is one in which there is a match in the vowel of the stressed syllable, and a difference in the consonants. In this study, I refer to the first form as consonance rhyme, and to the latter form as assonance rhyme. To find lines that contain repeated consonants, a method was developed which firstly selects the final stressed syllable. The initial consonants of this syllable are retained. In this syllable, all vowel sounds are replaced with a placeholder character, and the resulting pattern is used to find matches. To detect assonance rhyme, a text pattern is created in which all consonants are supplanted by placeholder characters. Obvi-

[412] *The Princeton Encyclopedia of Poetry and Poetics*, p. 1199.

ously, all instances of perfect rhyme needed to be removed before the start of the procedure.

Instances of assonance slant rhyme were found in the words “return” and “worth” in “Ode”, in “clover”, “sober” and “cream-soda” in “The Individualist Speaks”, and in “wisps” and “fist” in “Train to Dublin”. The algorithm also showed correctly that the well-known poem “Bagpipe Music” makes an almost exclusive use of feminine assonance rhyme (“python” and “bison”, “sofa” and “poker”, “whiskey” and “fifty”, “Blavatsky” and “taxi”, “Ceilidh” and “baby”). The use of slant rhyme effectively underlines the centre theme of the poem, as it focuses on the decline of a traditional rural culture of the Western islands of Scotland, and the hasty transition to a consumerist and urban culture. In this poem, the slant rhymes evoke a confused disorganisation, and suggest an abandonment of order and predictability. The insistent vowel sounds also mirror the lingering drone of the bagpipe. Consonance rhyme was detected in, among many other poems, “Birmingham”: “On shining lines the trams like vast sarcophagi move / Into the sky, plum after sunset, merging to duck's egg, barred with mauve” (ll. 25-26). It was also found in words such as “hourly”, “rarely” and “merely” in the poem “Entirely”and in “burden” and “garden” in “Prognosis”.

Instances of semi-rhyme were found by comparing the line endings that contain a single syllable to line endings that contain multiple syllables. When specific masculine rhymes match the stressed syllables in a feminine rhyme, this agreement is considered to be an example of semi-rhyme. Using this method, this form of rhyme was detected in lines that end in the words “eye” and “island” (“Street Scene”, ll. 32-33), “sky” and “horizon” (“Ode”, ll. 50-52), “way” and “neighbours” (“Eclogue from Iceland”, ll. 220-221), “skin” and “windows” (“Country Week-End”, ll. 84-86) and “run” and “Sunday” (“Sunday in the Park”, ll. 13-15). Evidently, this algorithm also returns line endings in which the stressed syllables are precisely the same, as in the final lines of the poem Mayflies (“May” and “mayflies”). It was decided, nevertheless, to count such cases as forms of semi-rhyme as well. The device is used frequently to emphasise the simultaneous sameness and diversity of phenomena. The poem “Wolves” contains a clear example. In this poem, the persona finds “pathos in dogs and undeveloped handwriting / And young girls doing their hair and all the castles of sand”. In this example, the matching acoustic patterns exemplify the inherent correspondences between the various phenomena which may shield the poet against the relentless unpredictabilities of reality. The poem “Snow Man” contains an equally functional use of semi-rhyme. The poem reflects on the nature of human memory and on the changing image of the self, by comparing it to a melting snow man. The second stanza muses on the decay of the remembered self-image: “Tomorrow / Comes the complete forgetting, the thaw / Or is it rather a dance of water”. In these lines, the solid shape of the single syllable dissolves, like melting snow, into a more fluid and more ephemeral state. Systematic manipulations of phonetic transcriptions can clearly help to demonstrate the verbal skills of MacNeice and his ability to create engaging acoustic effects.

### 6.2.4. Deibhide rhyme and internal rhyme

Robin Skelton has argued that MacNeice's Irish background manifested itself strongly in his frequent use of Celtic devices such as deibhide rhyme and aicill rhyme. Deibhide, firstly, is a form of rhyme in which a stressed syllable rhymes with an unstressed syllable. To trace occurrences of this first type of rhyme, an algorithm was developed which initially selected the final syllables of all lines within a stanza, regardless of whether these were stressed or unstressed. All initial consonants were removed. Lines were taken to contain deibhide rhyme if there was a match in the phoneme sequences of stressed and unstressed syllables, as in the following cases:

Holidays should be like this / Free from over-emphasis ("PostScript from Iceland", ll. 37-38)
The trains pass and the trains pass, chains of lighted windows (...) / For these are the trains in which one never goes ("A Contact", ll. 1-3).
But they all go so fast, bus after bus, day after day / Year after year, that you cannot mark any headway ("The Glacier", ll. 7-8)

During a review of the initial results, it was found that the algorithm disregarded a number of cases which seemed to contain obvious occurrences of deibhide rhyme. The poem "No More Sea", for instance, contains lines ending in "sum" and "medium". The final syllable of "medium" is transcribed phonetically using an unstressed schwa, and for this reason, it does not match the vowel in the word "sum". Problems such as these are difficult to solve. One option would be to work with a highly forgiving form of matching in which the schwa and the /ʌ/ vowel are viewed as identical, but this would result in matches on other locations which are clearly not intended as rhymes. To recognise noteworthy sonic patterns, the computer must be lenient in some cases, and strict in other situations. Capturing the appropriate level of flexibility clearly remains elusive.

Skelton also argues that the influence of the Celtic verse tradition is noticeable though the use of aicill rhyme. The *Princeton Encyclopaedia of Poetry and Poetics* explains that aicill rhyme consists of a correspondence between a phoneme sequence at the end on one line and a word in the interior of the line that follows immediately.[413] To study instances of this device, I examined windows of two lines. From the first line, the final stressed syllable was taken, which could optionally be followed by unstressed syllables. This pattern was compared to phoneme sequences in all the unique words of the following line. The last word of the second line was obviously neglected, as an agreement in the final words of two subsequent lines would constitute a regular perfect rhyme. It was also decided to disregard

[413] *The Princeton Encyclopedia of Poetry and Poetics*, p. 26.

lines that end in a single unstressed syllable. Among many other examples, the following instances could be detected:

Far more fatal than painted flesh or the lodestone of live hair / This despair of crystal brilliance ("Circe", ll. 4-5)
Filleted sun streaks the purple mist / Everything is kissed and reticulated with sun ("Morning Sun", ll. 8-9).
Stressing the function scrapping the Form in Itself / Taking the horse from the shelf and letting it gallop (Autumn Journal, section XII, ll. 50-51)

Internal rhyme, or occurrences of perfect rhyme within a single line, may be found by firstly compiling a list of all the unique words within a line, and by finding the phoneme sequences that occur more than once. This form of internal rhyme was found in the following fragments:

Its cage is a stage its perks are props ("Budgie", l. 4)
And the sun stood still above Notting Hill Gate ("Jericho", l. 26)
In quiet in diet in riot in dreams ("The Creditor", l. 4)

Skelton also discusses a specific form of aicill rhyme in which the final consonant sequence of one line reoccurs in the lines that follow. An attempt was also made to find this particular form of repetition. I implemented a procedure which analyses couplets of consecutive lines. All vowels are removed from the final words of the first line and from the entire second line. If the sequence of consonants is found to reoccur in the second line, this is captured as a case of consonantal aicill rhyme. During the initial tests, it became clear that there were many cases in which the last word of the first line and the first word of the second line were identical. Following Eagleton, who insists that rhyme entails "a unity of identity and difference",[414] it was decided to ignore all cases in which words are repeated verbatim and to focus solely on sonic affinities in distinct words. Next to the examples which were also discussed by Skelton, this form of aicill rhyme was found in the lines "Of the light in the dark of the muted voice of the turning wild / World yet calm in her storm gay in her ancient rocks" ("All Over Again", ll. 17-18) and "Your smile and chivvy your limbs through a maze of pearly / Pillars of ocean death--and yet you force your way" ("The Casualty", ll. 29-30)

### 6.2.5. Metre

In this case study, I also attempted to automate parts of the scansion. The stress patterns of the verse lines could be obtained directly from the phonetic transcrip-

[414] Terry Eagleton, *How to Read a Poem*, p. 132.

tions that were produced in this study, since these included data on the primary and the secondary stresses in each word. The stress patterns that were found were also classified. Methods were developed for the recognition of iambs, trochees, dactyls and anapests in trimeters, tetrameters, pentameters, hexameters and heptameters. Both regular verse lines and their catalectic or incomplete versions were considered. As I found that the algorithm initially resulted in very few examples of perfect iambic pentameters, I implemented a more lenient method, in which one of the feet in the pentameter was allowed to contain one supplementary unstressed syllable. Using this logic, iambic pentameters were found in the lines "To every question gives the same reply" in line 6 of "The Conscript" or "You cannot argue with the eyes or voice" in line 289 of "Eclogue from Iceland". The line "We are dying, Egypt, dying" ("The Sunlight on the Garden", l. 18) was found to be an example of a catalectic iambic trimeter, and "Neither sense nor conscience stirred" ("Jericho", l. 33") was classified as catalectic trochaic trimeter.

This method for the recognition of the metrical character of the various lines did not function flawlessly. An obvious difficulty is that the regular stress patterns of words may change within the context of a particular verse line. The poem The Sunlight on the Garden, for instance, is largely written in catalectic trimeters. The word "and", in the poem's second line, "hardens and grows cold", clearly receives stress within this rhythm, while it is listed as an unstressed word in the pronunciation dictionary. Complications such as these are difficult to solve, however, without a manual rearrangement of the stress patterns in the phonetic transcriptions.

### 6.2.6. Alliteration, assonance and consonance

In his essay "Feeling into Words, Nobel laureate Seamus Heaney has noted that the Ulster accent is "generally a staccato consonantal one".[415] Devices such as alliteration and consonance are correspondingly highly common among poems from Northern Ireland. In this study, I also attempted to analyse the ingenious repetitions of constants and of vowels within MacNeice's poems. Occurrences of alliteration were found, firstly, by counting the frequencies of all the consonants and the vowels that occur at the beginning of syllables that have primary or secondary stress. In this study, it was decided that when words are repeated, these repeated sounds do not constitute alliteration. For this reason, all repeated words were removed. I also assumed that alliteration takes place when sounds at the beginning of stressed syllables occur more than once. Following this logic, a heavy use of alliteration was found, for instance, in "Prayer Before Birth" ("With Strong drugs dope me with wise lies lure me, on black racks rack me in blood-baths roll me" / [...] / With water to dandle me grass to grow for me trees to talk" (ll. 6-9).

[415] Seamus Heaney, "Feeling into Words", in: *Finders Keepers*, London: Faber and Faber 2003.

Alliteration was also found in the second line of the poem Belfast: "Frozen into his blood from the fire in his basalt". This line is particularly interesting as the sounds that are repeated are the same as the main sounds in the name of the city that is depicted in the poem. Abrams suggests that the term 'alliteration' mainly refers to repetitions of consonants. Restricting the algorithm to occurrences of consonants, however, would result in inattentiveness to a number of striking instances, such as "Anger and ambush" in line 18 of "Iceland", "Who was innocent and integral once", in line 137 of "Ode" and "He plods the endless aisles not daring to close and eye" in line 2 of "Flower Show". It was decided, for this reason, to search for matches both in consonants and in vowels.

The method that was implemented in this study returned a pattern that represented the specific sounds that were repeated. On the basis of the complete list of patterns, it can be concluded that the poet deliberately planned alliterations. MacNeice frequently applied patterns in which particular pairs of sounds are repeated, or in which two alliterative words are nested within two other words whose initial consonants also alliterate. For lines 20 and 21 of "Western Landscape", for instance, ("And hanging smell of sweetest hay / Weavingly laughingly leavingly weepingly") the software produced the pattern "- h s - s h / w l l w". Examples of repetitions of pairs of sounds can be found in the lines "Prowl and plunge through glooms and gleams" (l. 20), and "A halfway house between sky and sea being of the water earthy" (l. 80) in "Donegal Triptych". Instances of alliterations nested within other alliterations can be found in the lines "Posed by Picasso beside an endless opaque sea", in "An Eclogue for Christmas", l. 33 or "In certain long low islets snouting towards the west", in "Last before America", l. 19. A similar, less noticeable, structure occurs in "Let us too make our time elastic" in "Mayfly", l. 23.

Similar methods were developed for the recognition of assonance and consonance. The *Princeton Encyclopedia* defines consonance as "the repetition of the sound of a final consonant or consonant cluster in stressed, unrhymed syllables near enough to be heard together".[416] Following this definition, I developed an algorithm which considers all possible windows of four consecutive stressed syllables in each verse line. I decided that the lines contains consonance or assonance if at least one of such clusters contains a repetition of a consonant or vowel at the end of these syllables. Using this logic, assonance was found in the lines "this dyspeptic age of ingrown cynics" ("Eclogue from Iceland", l. 117) and "In the sun-peppered meadow the shepherds are old" ("Nuts in May", l. 5). Similarly, consonance was found in the lines "Have seen myself sifted and splintered in broken facets" ("An Eclogue For Christmas", l. 34), "Metal patents parchment lampshades harsh" ("Belfast", l. 11), "The tight-lipped technocratic Conquistadores" ("Epitaph for Liberal Poets", l. 13).

[416] *The Princeton Encyclopedia of Poetry and Poetics*, p. 299.

### 6.2.7. Repetitions of words

Neil Corcoran has discussed MacNeice's use of refrains. It is a device in which a line, or a part of a line, is repeated verbatim within a poem.[417] Corcoran surmises that "repetition at the formal or technical level can be thematically functional",[418] and illustrates this statement via discussions of the poems "Leaving Barra", "Train to Dublin", "The Hebrides" and "Meeting Point". As part of the experimentation that was conducted for this thesis, I devised a method to identify repeated occurrences of n-grams of two or more words. Since refrains are taken to consist of literal repetitions, all function words, such as articles and pronouns, were retained during the creation of n-grams.

Corcoran's article T*he Same Again? Repetition and Refrain in Louis MacNeice* discusses 11 poems that contain refrains. The software that was developed for the case study was able to recognise the refrains in these same poems, and also identified a large number of additional examples. In the poem "April Manifesto", the word sequence "our april must replenish" occurs three times. The poem uses imagery of spring and of regeneration to represent the aesthetic delight that may be derived from an experience of colours and of sounds, and combines this with a mild criticism of consumerism. The repetition of the refrain at the start, middle and end of the poem effectively stresses the insistence of the craving for abundance. "An Eclogue for Christmas" contains several repetitions of the phrase "What will happen", and this underscores the poem's central sense of doom and anxiety about the future. In "Evening in Connecticut", the line "only the shadows growing longer and longer" is used both in the first and in the final stanza. At the beginning of the poem, these words are used mainly to describe the tranquillity of the evening and of the natural surroundings. The poem as a whole, however, mainly reflects on the immanence of a World War in Europe, and towards the end of the poem, the lengthening shadows essentially symbolise the darker qualities of human nature.

In some cases, the repetitions that were recognised by the application did not always constitute clear examples of refrains. In the poem "Invocation", for example, repetitions of the phrase "fetch me far" were found at the beginning of 14 lines in the poem. These repetitions, which collectively underscore the dream-like and escapist surge of the poem, ought to be classified more appropriately as anaphora. The same can be claimed for the repetitions of "It's no go" in "Bagpipe Music". In this poem, according to Neil Corcoran, the "almost demented repetitiveness propels the poem's hurdy-gurdy rhythmic relentlessness, as if the repeated phrase has taken over the poem".[419] If it was found that two or more lines contain the exact same words at the beginning of the line, such cases were captured as

[417] *The Princeton Encyclopedia of Poetry and Poetics*, p. 1151.
[418] Neil Corcoran, "The Same Again? Repetition and Refrain in Louis MacNeice", p. 216.
[419] Ibid.

instances as anaphora. Other examples of anaphora were found in "June Thunder" ("If only you would come [...] / If only now you would come", ll. 24-25), in "Plurality" ("Conscious of guilt and vast inadequacy [..] / Conscious of waste of labour", ll. 73-75) and in "Twelfth Night" ("O crunch of bull's-eyes in the mouth / O crunch of frost beneath the foot", ll. 2-3).

The software also pointed to an echo of "their verdure dare not show" at the beginning of "Valediction", but this phrase is repeated within a single verse line. Since it is stressed in the *Princeton Encyclopaedia of Poetry and Poetics* that refrains ought to be separated "by at least one line of nonrepeating material",[420] the application was enhanced with measures to ensure that the refrains do not occur on the exact same line. Furthermore, the simple fact that sequences of words are used more than once does not automatically imply that these are also meaningful refrains. The software identified recurrences of the words "here it was" in "A Hand of Snapshots", and of the phrase "over the wall" in The Stygian Banks. The reasons why these particular phrases should not be viewed as refrains, however, are difficult to codify in an algorithm. This finding stresses the continued need for manual evaluation and correction of the results.

In his discussion of MacNeice's "Leaving Barra", Corcoran also notes that the poem contains a specific form of repetition in which words or parts of words are repeated in other words within the same line. Examples can be found in the phrases "the rain and the rainbow" and "a belief that is unbelieving". Gillis argues that, in MacNeice's later writing, such forms of repetition "becomes a pivotal means of exploring emptiness and destabilization in the late Fifties". While Corcoran does not propose a term for this stylistic device, it may expediently be referred to as paronymy. This term refers to "two or more words partly identical in form and/or meaning, which may cause confusion in reception or production".[421] The recognition of this stylistic device can partly be automated. A first version of the algorithm simply established whether or not a word was contained in any of the other unique words on the same line. This method, however, was too crude, as it was found that words that consist of a single character, such as the article 'a', and the personal pronoun 'I', are obviously contained very frequently within other words. Since it was assumed that the device predominantly occurs in nouns, verbs and adjectives, a second version of the algorithm made use of the POS tagging, and initially removed prepositions, conjunctions and articles from all lines. This second method also overlooked a number of cases which did seem relevant, however. The lines "Beyond these plains' beyondless margin" and "Yet standing here and notwithstanding / Our severance" in "Letter from India", or "So much themselves in despite of spite" in "Visitations" were ignored, because of the removal of prepositions. A clear difficulty is that prepositions are relevant in some cases, but

[420] *The Princeton Encyclopedia of Poetry and Poetics*, p. 1151.
[421] R. R. K. Hartmann & Gregory James, *Dictionary of Lexicography* (Routledge 2002), p. 192.

irrelevant in other situations. The repetition of "in" on line 13 of "June Thunder", for instance, ("Then the curtains in my room blow suddenly inward") does not seem of significance.

An improved version of the algorithm made use of a list of stopwords. The standardised Glasgow list of stopwords, which is also employed in the Voyant software, proved to be too inclusive for this particular purpose.[422] Since words such as "sometime", "part", "together" and "give" are all on the Glasgow list, using this resource had the effect that striking formulations, such as "Next year is this year, sometime is next time, never is sometime" in line 76 of "Homage to Clichés", "Made him a part of the not to be parted whole" in line 71 of "Western Landscape", "Forgive what I give you" in line 1 of "To Mary", and "The more there are together, Togetherness recedes" in line 7 of "Babel" were all ignored. It was decided, for this reason, to make use of an edited list of stop words.

An important shortcoming was that the methods that have been discussed all failed to recognise one of the examples which was highlighted by Corcoran. The phrase "a belief that is disbelieving" is not recognised as paronymy, as "belief" is obviously not repeated in its entirety in "disbelieving". Making use of lemmatisation also would be ineffective in this particular case, as the root form of the latter word is "disbelieve". As a solution to this difficulty, I developed a method in which all possible substrings were extracted from words that contained more than three characters, excluding words on the edited list of stopwords. These substrings were used as the basis for the comparison. This procedure retrieved a number of noteworthy lines from the corpus, including "Of all desirable things - that is what I desire" in "Troll's Courtship", "Our past we know / But not its meaning — whether it meant well" and "Memories I had shelved peer at me from the shelf" in "Carrick Revisited", "what to these does the word significant signify" in "The Stygian Bank", "All the unconsummated consummations" in "The Return", "More than the twanging dazzle or the dazzling noise" in "Ode", and "the scalloped / Lampshade swings a shadow" in "Trilogy for X".

### 6.2.8. Onomatopoeia

Algorithms may similarly be created for the detection of onomatopoeia, or words whose phonetic aspects mimic the sound of the things they refer to.[423] The sounds of the words when pronounced could obviously be derived from the phonetic transcriptions, but data on the sounds produced by the referents of words were clearly unobtainable. There appear to be two distinct cases of onomatopoeia.

[422] Stop lists are resources which list the words which are most frequent within a particular language, and which, according to its developers, are of less importance for an analysis of the semantic contents of texts. The Glasgow list of stopwords was developed by The Information Retrieval Group at the University of Glasgow and contains 319 words.

[423] "Onomatopoeia", in Chris Baldick, *The Oxford Dictionary of Literary Terms*.

Firstly, specific words are onomatopoeic regardless of their context. Secondly, there are also words which assume onomatopoeic qualities only because of their use within a specific context. One example of the latter type of onomatopoeia can be found in "The Cyclist", which narrates a scene in which a boy cycles down a hill past the Westbury White Horse. The first line of the poem, "Freewheeling down the escarpment", effectively evokes the wind racing past the cyclist. While it seems virtually impossible to detect the latter type of onomatopoeia, the first type may potentially be detected by making use of a terminable list of words whose sounds resemble the acoustic aspects of the thing they represent. The list that was used for the purpose of the current case study also includes all inflected forms of verbs and both the singular and verbal forms of nouns, in order to ease processing. Instances of onomatopoeia were found in phrases such as "crickets fiddled and sizzled to drown the river" ("The Rest House", l. 5), "the squelch of mud the belch of surf" ("A Hand of Snapshots", l. 45) and "semaphore ultimatums tick by tick" ("The Hebrides", l. 108). Polysemous words such as "flush" or "spark" are clearly not onomatopoeic in all contexts. The lines "No spark of reality possible" ("Eclogue by a Five Barred Gate", l. 62) or "To keep it flush with the earth" ("Under the Mountain", l. 6) contain words which were included in the list of onomatopoeic words, but do not form compelling examples of onomatopoeia. As was the case for other figures of speech, the list of results that was produced by the software still needed to be examined and amended manually.

### 6.2.9. Allusions

During his lifetime, MacNeice produced an impressive body of literary criticism. *The Selected Literary Criticism of Louis MacNeice*, edited by Alan Heuser, includes essays on W.H. Auden, T.S. Eliot and Dylan Thomas, and following the death of W.B. Yeats in 1939, MacNeice also wrote an extensive critical study of the poetic works of his compatriot. It may be assumed that MacNeice's voracious reading also helped to shape the nature of his poetic development. Several authors have noted, for instance, that MacNeice's book on Yeats reveals as much about the older poet as about the author himself.

Among many other ways, MacNeice's literary influences may be explored by studying the explicit allusions to other literary texts. Crane argued that algorithmic explorations of allusions can be based on lexical similarity, on syntactical similarity or on phonetic agreements. In this case study, the attempt to identify allusions was based on agreements in vocabulary. An experiment was conducted which concentrated on the detection of the lexical parallels between poems by MacNeice and by W.B. Yeats. The algorithm that was developed firstly identified all the lines in the works of Yeats and of MacNeice which share two or more words. If there are clear lexical parallels in the usage of words, such correspondences may, in some cases, be characterised as allusions. To attenuate the impact of inflections, the algorithm made use of the lemmatised versions of the verse lines. In addition, all stop words

were removed, using the same manually edited list that was used for the detection of refrains. It was found, however, that this initial method resulted in an inordinately high number of combinations. To reduce the number of results, I added a requirement which stipulated that the shared lemmas should also be used in the same sequence. The number of verse lines that share such sequences of two lemmas was still decidedly high. Within the 16,782 lines by MacNeice and the 11,937 lines written by Yeats, I found 8,633 instances of lines with shared words. No matches of three or more words were found, however.

The algorithm assuredly helped to identify a number of compelling resemblances between poems by MacNeice and by Yeats. The opening line of the late poem "Flower Show", for instance, ("Marooned by night in a canvas cathedral under bare bulbs") contains an ironic invocation of the sixth part of Yeats's "Under Ben Bulben", in which the poet describes his epitaph ("Under bare Ben Bulben's head / In Drumcliff churchyard Yeats is laid", l. 91). The method that was developed confirmed Edna Longley's observation that the line "All you do is burke the other and terrible beauty" in "Eclogue by a Five Barred Gate" alludes to Yeats's line "A terrible beauty is born" from "Easter 1916". [424] The results produced by the algorithm also showed, interestingly, that the line that follows ("all you do / Is shear your sheep to stop your ears") (ll. 37-39) shares two consecutive lemmas with Yeats's "To a Shade". The latter poem consists of a reprimand against the Irish people who scorned Parnell, the founder of Irish Parliamentary Party. The poet advises the ghost of Parnell to "gather the Glasnevin coverlet / About your head till the dust stops your ear" (ll. 21-22). While "To a Shade" urges the ghost of Parnell to escape contemporary political realities, MacNeice's eclogue is conversely an exhortation to the shepherds to become more politically involved.

Many additional verbal parallels were found, some of which were significant. Common words were found, for instance, in "The Closing Album" and in "The Nineteenth Century and After". Part V of the former poem depicts the sentiment which is also expressed in the opening chapter of MacNeice's book on Yeats. MacNeice wrote that, after he had heard the news about the outbreak of the Second World War, this news established a new reality, which made nonsense of the old reality. "The Closing Albums" asks why the sea must continue to "draw a film of muslin down the sand / With each receding wave?". A parallel was found in Yeats's brief poem "The Nineteenth Century and After", which similarly focuses on the end of a period. Yeats urges readers, nevertheless, to appreciate the present and to value the "rattle of pebbles on the shore / Under the receding wave". Similarly, the line "That Man is a dancer is an anachronism" in "Precursors", may be viewed as an allusion to Yeats's "Nineteen Hundred and Nineteen", which asserts that "All men are dancers". The phrase in the Yeats poem is used in connection with the "Platonic Year", which, according to Jeffares, represents the notion that "the whole of the

[424] Edna Longley, *Louis MacNeice: A Study* (London: Faber and Faber 1988), p. 102.

constellation returns to the positions from which they once began".[425] MacNeice's poem suggests that such cultural and ideological renewal can no longer be achieved in the late twentieth century. The "topless tower" in "Brother Fire" may be viewed as a reference to the "topless towers / Where Helen walked with her boy" which are depicted in Yeats's "When Helen Lived". Both poems focus on a destructive force within human nature, which obstruct a full commitment to beauty in times of hardship. Section XIV of *Autumn Journal* depicts a journey through rural Oxfordshire, and on line 11, the car's windscreen wiper is compared to a "cricket that sings". This image clearly echoes the line "Dropping from the veils of morning to where the cricket sings" in "The Lake Isle of Innisfree". Both poems, notably, concentrate on the redemption that may emanate from a retreat into nature. "Old Masters Abroad" likewise contains a references to "The Lake Isle of Innisfree" in the line "Nine bean rows rise in the Kalahari".

As was shown, a method which connects lines that share two or more consecutive lemmas can disclose noteworthy similarities between the works of different authors. Whereas such lexical parallels do not always indicate intended allusions, placing such different literary contexts side by side can occasionally quicken reflection on similarities or contrasts. A first problem with the method that was followed, however, is that it returns a high number of shared bigrams, while only a minority of these cases seem significant from a critical point of view. Unfortunately, as noted by Coffee et al., the precise quality that renders an allusion relevant or significant often remains incommensurable.[426] The phrase "We are dying, Egypt dying" in "The Sunlight on the Garden" is an obvious allusion to Shakespeare's "Anthony and Cleopatra", but the algorithm that was implemented in this experiment also singled out lines containing common collocations such as "some day", "close eye", "so long", "young man", "come from", "each other", "all day" and "far away". Moreover, intertextual reference are frequently highly complex, and a method crudely based on shared bigrams callously misses such more intricate types of allusion. The opening line of the second section of "Autumn Journal", "Spider, spider, twisting tight" is an obvious allusion to William Blake's "Tiger", but a method based on verbal similarities would evidently fail to link the two lines. The first stanza of "Brother Fire" depicts the fires during the London Blitz as a dog raging through the streets, and this imagery is reminiscent of the phrase "let slip the dogs of war" from Shakespeare's "Julius Caesar". The recognition of such extended allusions clearly demands a more sophisticated matching algorithm. Richard Danson Brown also notes that the third stanza of "Neutrality" echoes "the vocabulary and idiom" of Yeats and that the phrase "Intricacies of gloom and glint" is reminiscent of the line "In all lovely intricacies of a house" from

[425] W.B. Yeats, Norman Jeffares (ed.), *Yeats's Poems* (Basingstoke: Macmillan 1996), p. 583.
[426] Neil Coffee et al., "Modelling the Interpretation of Literary Allusion with Machine Learning Techniques".

Yeats's "In Memory of Major Robert Gregory".[427] Brown suggests that allusion may also be based on a singular common word. The precise reasons why repeated words constitute allusion in some cases, while they do not in other situations, seem impervious to formalisation. A method based on shared n-grams may help to identify a number of striking parallels, but the task to separate the significant matches from the insignificant matches is highly labour-intensive. In addition, the method also missed many more perplexing categories of intertextual reference. As an idea can be expressed in many different ways, using very different words, a method that is based on ngrams only has limited value. Because of these difficulties, the attempt to recognise literary allusions automatically has not been pursued further.

### 6.2.10. Imagery

In his monograph *Sceptical Vision*, Terence Brown devotes a full chapter to the discussion of MacNeice's imagery. One of the images that Brown concentrates on is that of the sea, which, as he argues, features in many poems as "a major image of eternity, of the beyond, of non-being".[428] In "Western Landscape" and "Around the Corner", conversely, the sea is mostly portrayed as a redeeming force.[429] Brown adds that other natural elements, such as wind and stones, frequently have a particular significance as well. Compelling examples of references to stones and to petrification can be found in "Nocturne", "The Glacier" and "Western Landscape". Brown also discusses the poet's references to various modes of transportation, such as cars, boats and trains. In the poem "Trilogy for X" the train represents the notion that the persona moves "through a world of vanishing particulars where new data or new phenomena present themselves continually".[430] A related image is that of the quest, which often represents the poet's uneasiness with an ever-changing reality and the central hankering after profundity.

Computation may enable scholars to trace references to specific images, but a computer-based analysis of imagery is complicated by a degree of opacity with respect to the precise definition of the term. Baldick notes that imagery is a "rather vague critical term covering those uses of language in a literary work that evoke sense-impressions by literal or figurative reference to perceptible or 'concrete' objects, scenes, actions, or states".[431] In his essay "Modern Poetry", MacNeice also makes an important distinction between, on the one hand, the properties of a poem, which are essentially the object and the actions which are needed to construct a narrative, and, on the other hand, the images proper, which are words

[427] Richard Danson Brown, "Neutrality and Commitment: MacNeice, Yeats, Ireland and the Second World War", in: *Journal of Modern Literature*, 28:3 (2005), p. 114.
[428] Terence Brown, *Louis MacNeice: Sceptical Vision*, p. 118.
[429] Ibid., p. 121.
[430] Ibid., p. 106.
[431] "Imagery", in Chris Baldick, *The Oxford Dictionary of Literary Terms*.

which primarily have a metaphorical function and which represent more recondite concepts.[432] As it seems unattainable to make a distinction between properties and images on formal grounds, however, these two types of images will be treated equally.

In this case study, a first attempt to extract imagery from the corpus was based on topic modelling. In this experiment, the MALLET program has been set to recognise 100 topics. An evaluation of the results of MALLET revealed that the words which occur frequently in the same documents were placed in one topic. None of the word clusters that were produced, however, could easily be resolved to a recognisable image. As was also demonstrated by other authors,[433] topic modelling and LDA do not seem helpful for the study of imagery in corpora containing figurative language.

In this study, attempts to identify imagery algorithmically were also based on the UCREL Semantic Analysis System (USAS) and on The Harvard General Inquirer (HGI). USAS, firstly, as was discussed in Chapter 3, is a semantic tagger which can connect the concrete tokens found in a document to broader semantic categories.[434] The entire MacNeice corpus was tagged using USAS, and categories that were applied most frequently included A3 ("Being"), B1 ("Anatomy and physiology"), M6 ("Location and direction"), M1 ("Moving, coming and going"), O2 ("Objects generally"), O4 ("Physical attributes"), T1.3 ("Time:Period"), L3 ("Plants").[435] Categories which refer to abstract concepts or to specific components of argumentative disposition, such as the categories in section A1 ("General and abstract terms") and Z ("Names and grammatical terms") were largely ignored, since this study is interested in words that that refer to concrete objects or events. The initial analysis offers some support for Brown's observation that MacNeice was preoccupied with movement, travelling and time. The USAS classifications also exposed references to types of imagery that was not discussed in the criticism that was surveyed for this study. It was found, for instance, that poems also contain many references to plants and to flowers: "The murderous grin of toothy flowers" ("Intimations of Mortality, l. 14), "Coral azalea and scarlet rhododendron" (Ode, l. 99), "There is more than glass between the snow and the huge roses" ("Snow", l. 12), "Frost will not touch the hedge of fuchsias" ("Valediction, l. 76), "a welter of nasturtium" (The Closing Album, l. 89). One difficulty with USAS is that the coverage of its categories is decidedly broad for this specific purpose. The USAS categories do not correspond directly to the imagery that is discussed by earlier critics of MacNeice's verse. Browns focuses on highly specific images, such as that of the sea, storms or church bells, but, within USAS, such phenomena or objects

[432] Louis MacNeice, *Modern Poetry: A Personal Essay*, p. 91.
[433] Lisa M. Rhody, "Topic Modeling and Figurative Language".
[434] A guest account to WMatric and to USAS was kindly provided by dr. Paul Rayson, Reader in Computer Science at the University of Lancaster.
[435] <http://ucrel.lancs.ac.uk/usas/semtags.txt> (18 July 2014))

are mostly subsumed within broader terms. Terms pertaining to quest imagery, for example, partly correspond to the lexicons for categories such as M1 ("Moving, coming and going"), M2 ("Putting, taking, pulling, pushing, transporting &c. "), M3 ("Movement/transportation: land") and M4 ("Movement/transportation: water "), but these headings also cover many terms that are clearly unrelated to the notion of the quest.

Imagery may also be detected, to some extent, by making use of the HGI, which is "a computer-assisted approach for content analyses of textual data".[436] The lexicons can be used to identify words with a negative or a positive connotation, words which express strength or weakness, or words which indicate a passive or an active orientation, among many other categories. A minority of lexicons contain terms that refer to more concrete phenomena or objects. One subclass enumerates words that describe "created locations that typically provide for social interaction and occupy limited space".[437] It includes lists for tools, types of food, vehicles and buildings. Furthermore, the HGI also provides a list of words denoting colours. The tool can usefully be applied to recognise tokens in the general categories which were defined within HGI. As is the case for USAS, however, the HGI operates on semantic fields which are generally much broader than those which interest literary critics. One of the images which are discussed by Terence Brown, for instance, is that of the Madonna. The image figures prominently in poems such as "Belfast" and "Evening Indoors". Within HGI, by contrast, the term "madonna" is subsumed under the much broader category "religion".

In this study, a solution was developed in which a number of customised lexicons were developed, using the lexicons that were supplied by USAS and by the General Inquirer as a basis. All categories from the USAS system which do not refer to concrete events of objects, such as A13.1 ("Degree: Non-specific") or N5 ("Quantities") were removed. Lexicons from USAS and from the General Inquirer which refer to similar phenomena, such as "Religion" and "religion and the supernatural" were merged. Broad categories in the sections "M" and "T" were replaced manually with narrower categories, so that searches for specific images such water, wind, stones, threads also became possible. These bespoke lexicons, which cluster words that evoke specific objects, events or sensory impressions,[438] helped to identify many of the images which are used in MacNeice's poetry.

To gauge the accuracy of this method, all poems which are discussed in the chapter "The Poet and his Imagery" from Brown's monograph on MacNeice were listed, in combination with the images these contain. Additionally, an application was developed in which the results of the method based on machine reading were

[436] <http://www.wjh.harvard.edu/~inquirer> (21 July 2014)
[437] <http://www.wjh.harvard.edu/~inquirer/homecat.htm> (21 July 2014)
[438] The method that was used in the case study of this thesis is similar to the method that is described in John B. Smith, "Computer Criticism", p. 20.

compared with the results of Brown's readings of the poems. This comparison showed that the software managed to identify the same images as Brown in the majority of cases. This was the case for poems which, according to Brown, contain references to threads, church bells, the sea, trains and boats. Additionally, the software pointed to additional references to these same images in many other poems. It identified 40 poems containing thread imagery which were not listed in Brown's study. Brown explains that the image of a thread or a ball of wool often represents "life in perpetual change in time"[439], but it is also used to describe the permanence of the ancestral background in "Valediction": "The woven figure cannot undo its thread" ("Valediction", l. 43). In "Postscript to Iceland", by contrast, the line "All the wires are cut, my friends" (l. 67) indicates a separation from past securities. The software for the detection of imagery also pointed towards 43 additional poems which confirmed Brown's observation that church bells often have a sinister connotation. It has this effect in "Sunday Morning" ("the church spire / Open its eight bells out skulls' mouths which will not tire", ll. 10-11), "Half Truth from Cape Town" ("Between a smoking fire and a tolling bell", l. 1) and section XVI of "Autumn Journal" ("yet her name keeps ringing like a bell", l. 59). The results of the algorithm were not equally convincing for the poems in which Brown found quest imagery, however. Many of these poems conjure up the image of the quest through related concepts such as voyages, lures and forms of hankering. Such concepts are not evoked systematically though a fixed set of terms. An additional complication is caused by the fact that the precise meanings of strings are often determined situationally. While the recognition of imagery functions well in most cases, the manifold semanticity of words occasionally resulted in obvious errors. The phrase "wind your gramophone", in line 22 of "An Eclogue for Christmas" was labelled as an instance of wind imagery, and the phrase "one rocks a firelit / Cradle", in lines 116 and 177 of "Flowers in the Interval" was identified as an example of rock imagery. Next to these incidental errors, which were removed manually, the software retrieved relevant references to images in the majority of cases.

### 6.2.11. Themes

Poems generally epitomise certain broad ideas or general emotions. Whereas the direct denotation of the words in the text can sometimes be found by making use of semantic lexicons, there is rarely a logical relationship between the tokens and the deeper themes which are being developed by these tokens. Themes can, in most cases, be found exclusively via close reading and by being sensitive to the concepts and the emotions which are evoked. In this study, no attempts were made to identify themes algorithmically. Since it was estimated, nevertheless, that it was

[439] Terence Brown, *Louis MacNeice: Sceptical Vision*, p. 108.

necessary to have data about the thematic concerns of the poems, these data were supplied manually. In *Sceptical Vision*, Brown devotes three chapters to discussions of themes. In this study, themes were correspondingly divided into three main categories: (1) romanticism, (2) a rejection of modernity and (3) a concern with metaphysics. Under these three central headings, Brown also discusses more specific themes. Using Brown's broad classification, 15 specific themes were defined. Brown's texts also contained 152 assignments of poems to these themes. The remaining poems have been classified on the basis of close readings. Assigning themes to texts is obviously a subjective task, and it is also labour-intensive. Although computation can, in some cases, accelerate research or help to make the analyses more objective, this is clearly not the case for all aspects of computer-based research.

## 6.3. Conclusion

In his autobiography *The Strings are False*, MacNeice explains that he frequently felt divided between two contrastive philosophical impulses. While he "wanted the world to be One, to be permanent", he also acknowledged that "any typical monistic system [i.e. based on Oneness] appeared hopelessly static".[440] This conflict between a desire to understand reality through a uniform set of principles on the one hand, and a realisation of the perplexing and diversified nature of actual phenomena on the other also forms a central concern in MacNeice's verse.[441] Poems which express a longing for permanence and for consistency, such as *Western Landscape* or *Plurality*, often contain a simultaneous acknowledgement of the notion that the world, as formulated in the well-known poem "Snow", is "incorrigibly plural". In his poem *Entirely*, MacNeice presents the world as "a mad weir of tigerish waters" (l. 19) and writes that "when we try to eavesdrop on the great / Presences it is rarely / That by a stroke of luck we can appropriate / Even a phrase entirely" (ll. 5-8). Similarly, "Train to Dublin" celebrates the diversity and the discontinuity of the phenomena that appear before the persona's mind, while simultaneously recognising an inability to recognise any structure or consistency as the poet's "half-thought thoughts divide in sifted wisps / Against the basic facts repatterned without pause" (ll. 1-2).

The pursuit of mechanisms for the computer-based detection of specific qualities of literary texts is marked by a very similar tension. Algorithms consist of a finite collocation of univocal instructions, and can consequently anticipate a limited number of cases and exceptions only. Nevertheless, they have been used in this chapter to generate metrics about literary devices whose concrete manifestations

[440] Louis MacNeice, *The Strings Are False: An Unfinished Autobiography* (Oxford: Oxford University Press 1966), p. 125.
[441] Edna Longley, *Louis MacNeice: A Study*, p. 143.

often vary deeply. MacNeice's manifold proficient experiments with repetitions of sounds or of words, for instance, have patently resulted in an expansive variety. Particular phoneme sequences are repeated entirely or in part, and they are repeated both within lines and between different lines. There are incidentally verbatim repetitions of phrases, but, as also noted by Neil Corcoran, there are also many repetitions with variations.[442] It seems virtually impossible to formulate definitive rules for the recognition of linguistic repetition in its full plurality.

Next to the challenges arising from the fact that literary phenomena may display a sheer boundless variation, this chapter has identified three additional obstacles. Firstly, it was found that there is a class of textual features whose rule-based recognition, to a large extent, remains elusive. This is the case, for instance, for the identification of allusions. It can be stipulated, for instance, that allusion is likely to occur in lines that share two or more lemmas, but it remains difficult to predict when such correspondences are of actual literary significance.[443] This assessment, in most cases, can only be made by human scholars. A second difficulty is that software often supports binary distinctions only. It assumes that a feature is either present or absent. Certain cases are clearly ambiguous, and there can be good reasons both for accepting and for rejecting a specific result. The implementation of an algorithm implies the statement of a definition of the aspect to be detected, in the very literal sense of drawing a boundary between cases which are relevant and cases which are not. It is generally difficult to create programs that can qualify statements, or that can add nuances to results. While computers necessarily reduce options to either '0' or '1', or to either black or white, research in the humanities is often intent on exploring the many shades of grey that exist between such stark opposites.

A third and obvious difficulty is that computers can only process data which are available in an explicit form, or which can be inferred unambiguously from other data which are present explicitly. As a result of this limitation, there are numerous literary devices whose unsupervised identification remains difficult. In this thesis, it was estimated, for instance, that it was impossible to extract data about themes via an algorithmic approach. Other examples include metaphor, personification, synaesthesia, chiasmus and understatement. In most of these problematic cases, recognition demands an understanding of the semantic contexts of these devices. In literary writing, the semantic context is generally too complex and too unpredictable for current semantic taggers, and, in most cases, data which require an understanding of the text's meaning can only be supplied manually.

Algorithms for the recognition of literary devices almost inevitably incur error margins. McCarty notes that imperfection is inherent to all applications in the digi-

[442] Neil Corcoran, "The Same Again? Repetition and Refrain in Louis MacNeice".
[443] Neil Coffee et al., "Modelling the Interpretation of Literary Allusion with Machine Learning Techniques".

tal humanities, as the products of the human imagination can never be restrained or tamed entirely by a single conclusive algorithm. The use of computer-based methods implies a "continual process of coming to know by manipulating things, not an achievement but an approximating convergence".[444] Tools, more specifically, can be imperfect in three ways. Software can create a data set which is marked by a high degree of recall, but a low degree of precision. This means that the algorithms return a high number or results, many of which, unfortunately, are irrelevant. Secondly, there may be a high degree of precision, and a low degree of recall. This implies that, while the results that are returned are mostly correct, there also a large number of relevant fragments in the corpus which were not identified correctly. In the least desirable and third scenario, the software scores low both on precision and on recall. If it is indeed inevitable that tools for the identification of literary devices are structurally capricious and unreliable, a hybrid solution remains necessary, in which the computer initially produces data which subsequently need to be verified and, potentially, corrected by human scholars. This obviously places a certain limit to the scale of data sets. In the case of a low precision, scholars are confronted with a data set that contains a high degree of noise, and efforts will need to be taken to clean the results. In the case of a low recall, however, scholars will need to revisit the original sources and attempt to add data about the cases which the software had overlooked. As the removal of unwanted elements from a data set seems less labour-intensive than revisiting an entire corpus, casting the net widely and optimising recall generally seems preferable.

Terms such as 'perfection' and 'imperfection' are obviously subjective. The question whether or not an algorithm functions in a satisfactory manner can only be answered by relating its results to the expectations of individual researchers. Tools are invariably based on a prior assessment of likely forms of usage. The heuristic methods that are implemented are likewise based on assumptions and decisions that can mostly be contested. Even if an application is flawless according to one scholar's criteria, the tool may still be inadequate for scholars with different needs. All descriptions of textual aspects require at least some level of interpretation. The algorithms that have been proposed in this chapter have evolved in sequences of trial and error. They have been calibrated and tweaked on the basis a specific collection of texts, and it is highly probable that new imperfections and new inconsistencies are exposed when they are applied to other text collections.

Despite the almost inescapable shortcomings, algorithmic processing clearly produce a number of benefits. Without digital instruments, the recognition of distinct devices such as alliteration, consonance and internal rhyme would depend fully on the alertness of individual scholars. Verse generally contains a wide range of literary devices, and it is often difficult for scholars to be alert to all possible

[444] Willard McCarty, *Humanities Computing*, p. 28.

devices simultaneously. If the detection of devices can be caught in an algorithm, scholars can produce a more encompassing description of the devices that occur in the poetry.

Digital methods are often applied to allow for explorations of corpora in their entirety. The possibility to study the whole, nevertheless, depends critically on the availability of consistent and reliable data about individual parts. Studies which are largely based on word frequencies can mostly progress directly to the level of corpora, since algorithms for the segmentation of words have been tested extensively. If data are to be created by algorithms which are still experimental, however, analyses of data sets on a macro-level are generally preceded by an evaluation of individual data values on a micro-level. The software under development generates lists of results which are potentially relevant, and the data sets that result from potentially defective algorithms mostly require further editing. This process of editing, and the evaluations of the text fragments which were selected by the software, often lead in themselves to a better understanding of the texts that are studied. Enumerations of occurrences of literary devices are comparable to a concordance. Lists of devices, taken out of their original context, can enable scholars to investigate the ways in which these terms were used throughout a body of literary works.

Scholars who are involved in the development of software for the analysis of texts can generally enhance their understanding both of the texts that are studied and of the methodology used for studying these texts. The application of algorithms, and the subsequent revision of insights, based on the results produced by these algorithms, may be regarded as a form of dialogue between reader and text, with the digital tool as an intermediary. As such, digital scholarship effectuates a redress of what Plato viewed as a crucial deficiency of written text. One of his main reasons for denouncing the written word was that it halted any discussions, as it precluded the possibility of interaction. Once a text was solidified on a static surface, the argument could no longer develop.[445] Arguably, an algorithmic engagement with the text rekindles the text's capacity to respond. The digital tool may be seen as a rendition of the scholar's understanding or conceptualisation of the source, and, provided that no coding errors have been made, running an algorithm is in effect a scrutiny of the theoretical assumptions made during the algorithmic design. The data which are produced as output may prompt programmers to reconsider these assumptions and to reassess the algorithm as implemented. Through such cycles of iterative development, scholars can theorise through practical work.

Human scholars and computers both have strengths and limitations, and these manifest themselves in opposite ways. Human scholars can be fully attentive to individual cases and to complexities discernible in particular cases. Such a focus on

[445] Plato, *Phaedrus* (Cambridge: Harvard University Press 1953).

details, however, generally hinders a comprehensive assessment of large text corpora. The approach that is taken by the machine is, in many respects, the mirror image of the method followed by the human scholar. Algorithms can collect data about the corpus in its entirety, but their observations can often be inaccurate. A systematic exploration of a text corpus mostly requires a close cooperation between the machine and the human researcher, and close reading and algorithmic processing are best viewed as complementary methods. Scholars critically face the challenge to apply computation in a dexterous manner, and to strike an astute and heedful balance between close reading and machine reading, between accuracy and completeness, and between monism and plurality.