Tampere University

Santeri Joensuu

MĀORI INFLUENCE ON NEW ZEALAND ENGLISH
A corpus study on the different senses of kiwi

ABSTRACT

This bachelor's thesis is a brief look into the Maori word kiwi and how it is used in modern New Zealand English regarding the two senses it has. These two senses are of a small bird native to New Zealand and a New Zealander. The area of study is similar with Alexander Onysko and Andreea Calude's 2014 study of Maori words Maori, Pakeha and Kiwi but with a smaller scope, data set and a different focus. The thesis is a corpus study conducted using The New Zealand component of the International Corpus of English (ICE-NZ) and The Wellington Corpus. The findings in the thesis indicate a significantly stronger presence of the New Zealander sense. There is a distortion in the ICE-NZ data which makes one sense appear more common than it is, but this is accounted for in the thesis.

Keywords: corpus study, New Zealand English, kiwi, Maori

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Table of contents

1. Introduction

The relationship between New Zealand English and *te reo Maori* (the Maori language) is somewhat unique in the English-speaking world. Since the development of NZE in the 1800s, Maori words have been introduced into the lexicon and are used alongside traditional English words seamlessly. Other dialects of English do obviously have loanwords, but none to the same extent as NZE; this is the uniqueness of the dialect.

Origins of the NZE dialect has been studied before, notably by Peter Trudgill who offered his own theory on the formation on the dialect, including the borrowing of Maori words. Trudgill notes the significance of Maori as many words of Maori origin are contemporary in modern NZE and used by most of the population (2004, 4). Trudgill's extensive theory and study, however, does not understandably focus on single words.

There have been other studies where a smaller set of words has been the main subject. Alexander Onysko and Andreea Calude have studied three major Maori words, *Maori, Pakeha* and *Kiwi* – the native people, European New Zealanders, and just New Zealanders, respectively. Onysko and Calude note that there are previous hypotheses on Maori words in NZE, but these have not been analysed or confirmed closely (2014, 143), which is the reason for the existence of this thesis. The Onysko and Calude study namely focuses on differences in the modern usage of words; demographical differences and spoken-written differences. What is different in this thesis is that the focus is on the meaning behind the word *kiwi* instead of its total appearance in language.

The word *kiwi* is of Maori origin with the original meaning according to the Oxford English Dictionary being: "Any of several dark brown or grey tailless, flightless, nocturnal ratite birds of the genus *Apteryx*, of New Zealand forest and scrub, having a long bill and hairlike feathers. Cf. apteryx n." (23.11., 13.40). The word is one of the oldest Maori loans in NZE along with other flora and fauna words of Maori origin (Hay et al., 2008, 68). As times changed, *kiwi*

obtained another meaning: "(With capital initial.) A New Zealander, esp. a New Zealand soldier; (also) a New Zealand sportsman. Also attributive." (OED).

This new sense of the word is the motivation for this thesis. Can we determine the significance of this newer meaning of the word in comparison to the original meaning? And if so, would it be sensible in the future to observe the historical development of this difference? The expected findings from the data in this thesis is that the New Zealander sense is more common in contemporary language than the bird sense. This is purely based on the logic that New Zealanders speak of themselves significantly more than of birds.

2. Methodology

The source for the corpus data used in my study was the New Zealand component of the International Corpus of English (ICE-NZ) and The Wellington Corpus, both spoken and written versions. The reasoning behind this choice is the availability of the data through Tampere University. The optimal choice in the study would have been to use the ONZE data, a collection of recordings that includes the whole development of the NZE dialect which is something that is not available for any other English dialect or perhaps even any other language. The problem with that data is that it requires presence in the University of Canterbury in New Zealand and that would be more time consuming that is viable for a bachelor's thesis.

The corpus resource site of Victoria University of Wellington informs us that the ICE-NZ corpus was compiled by Victoria University of Wellington in 1999. It contains spoken and written language data from 1989 to 1994 in 600,000 and 400,000 words of speech and written language, respectively. The scope of the corpus is somewhat limited; compared to other corpora, for example the British National Corpus (which is very vast for a corpus), ICE-NZ is small. However, its contents are suitable for this thesis and quality compensates for quantity, which is

helpful as many corpora can include indecipherable data. The Wellington Corpus is chosen specifically to compensate for the small scope of ICE-NZ.

According to the corpus resource site of Victoria University of Wellington, The Wellington Corpus' two versions, spoken and written, were compiled in slightly different periods. The spoken version, WSC, was collected in 1988 to 1994 and contains 2,000 unique word extracts, amounting to roughly 1.5 million words. Victoria University of Wellington notes that as informal conversation data is difficult to obtain, the WSC and ICE-NZ have overlap in nine categories and 339,530 words. The written version of Wellington Corpus, WWC, was compiled from published writings from 1986 to 1990 and it contains 2,000 different word excerpts on several topics translating to approximately 1.1 million words. The Wellington Corpus is more extensive than ICE-NZ and gives significantly more data for this study. The exact number of word tokens found in all the data calculated by AntConc is 4,575,103.

Simple searches were the method of data collection for this study. To observe the presence of different senses of *kiwi*, several smaller searches using the AntConc programme were the chosen method. Each corpus data package was studied separately; the division was ICE-NZ spoken and written and Wellington spoken and written. This study includes two searches for each data package: 'kiwi' for the singular form and 'kiwis' for the plural form. The plural has the suffix -*s* while other Maori-origin forms do not have a plural form in NZE. *Kiwi* is an exception to this (Hay et al., 2004, 74). Due to the time constraints and the possible scope, this study contains no other searches. After the searches, I inserted the data into Excel charts and calculated the number of each sense's presence. The following is the division of senses used:

1) Kiwi as New Zealander

2) Kiwi as the bird

3) Kiwi as some other sense (e.g. referencing the New Zealand Dollar, 'kiwi dollar')

4) the sense was not clear from the data, i.e. ambiguous

The third category includes the attributive variant of the New Zealander sense as this thesis is more concerned with the differences in noun presence comparison of the first two categories. The data led to the formation of several different diagrams. For both corpora, the charts for singular and plural uses have been separated and then combined into one chart which have been further combined in a chart representing all the data in one.

3. Analysis of the data

In this section we will go through the data and results found in the study. First, the corpora are separated into two subsections and then we will look at the combined data at the end of the section and discuss the results further.

3.1. The New Zealand component of the International Corpus of English

The ICE-NZ turned out to be challenging to examine for this study. The number of tokens of the word 'kiwi' or is 129 for the singular variant.

	ICE-NZ Spoken	ICE-NZ Written	ICE-NZ Total	
Kiwi = New Zealander	35	17	52	
Kiwi = a type of bird	0	54	54	Total token count:
Kiwi = other sense	10	7	17	129
Ambiguous result	5	1	6	

Table 1. 'Kiwi' singular sense in ICE-NZ

As we can see from Table 1, the bird sense of *kiwi* is not at all present in the spoken data of ICE-NZ while the New Zealander sense is the most common, with 35 tokens. Unexpectedly other senses for *kiwi* were better represented. These senses were either 'kiwi card' or the New Zealand dollar, 'kiwi dollar'. Five of the tokens were ambiguous or the data surrounding the token was lacking, leading to failure to recognize the sense.

The problems begin with the written data. The New Zealander sense is less common in writing with 17 tokens, but the bird sense is significantly more present in the written data with 54 tokens. There is an explanation for what seems to be a mistake. All 54 of the tokens are from the same source in the corpus. With the size of ICE-NZ being small, this results in a significant distortion in the data and makes the analysis more difficult.
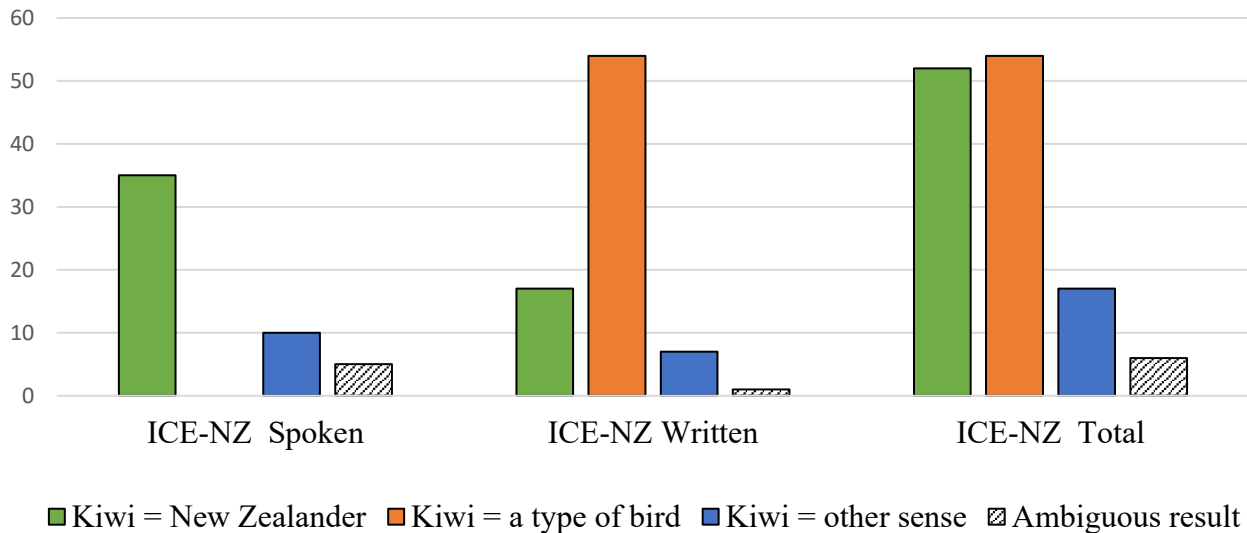

Diagram 1. 'Kiwi' singular sense in ICE-NZ

The significance of the distortion can be seen in Diagram 1. With one source having a large quantity of tokens, and in this study more than the other main sense in total, analysis is practically impossible as no other source gives us any data on the bird sense of *kiwi*.

We can observe the same problem with the plural to a higher extent. The total number of plural tokens for *kiwis* is 96. Again, in the spoken data the bird sense lacked any presence while the New Zealander sense had 41 tokens at the same time. The written data then is the opposite: just one token for the New Zealander sense while the bird sense has 50 tokens, all again from the same source text.
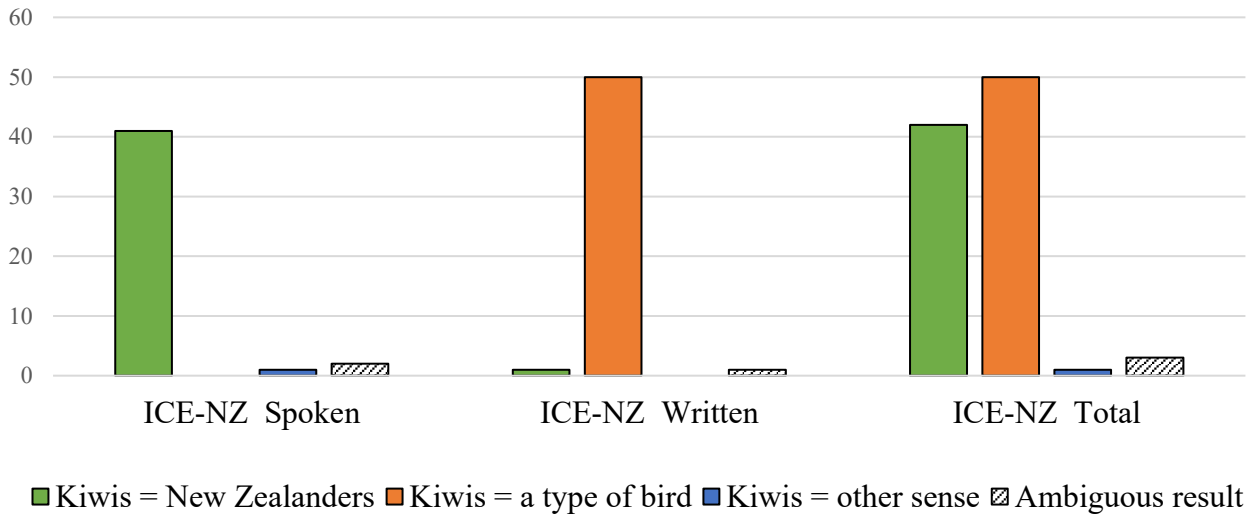
Diagram 2. 'Kiwis' plural sense in ICE-NZ

Diagram 2 shows us the more visible effect this distortion has on the plural data. The other senses are of no interest in this diagram as they are so minute. When compared to Diagram 1, the difference in written data is even more problematic as the data from the one source surpasses other data by an even larger margin.

	ICE-NZ Spoken	ICE-NZ Written	ICE-NZ Total	
Kiwis = New Zealander	76	18	94	
Kiwis = a type of bird	0	104	104	Total token count:
Kiwis = other sense	11	7	18	225
Ambiguous result	7	2	9	

Table 2. 'Kiwi' singular and plural combined in ICE-NZ

Looking at the combined data in Table 2 we can see the data is largely unusable as the one written source includes approximately 46,22% of the 'kiwi/-s' tokens in ICE-NZ. The data will still be used as the other senses of *kiwi* have usable data in the corpus. The distortion was unexpected, and it will show in the later combined data of ICE-NZ and Wellington Corpus.

3.2. The Wellington Corpus

Overall, Wellington data was more akin to the expectations for this study. As with ICE-NZ first thing to examine is the singular 'kiwi'. In both the spoken and written data combined the total number of tokens for 'kiwi' is 120.

	Wellington Spoken	Wellington Written	Wellington Total	
Kiwi = New Zealander	42	32	74	
Kiwi = a type of bird	0	3	3	Total token count:
Kiwi = other sense	16	23	39	120
Ambiguous result	3	1	4	

Table 3. 'Kiwi' singular in Wellington

As with the ICE-NZ data, Table 3 shows us the dominance of the New Zealander sense in spoken contexts with 42 tokens compared to the bird senses zero tokens. Interestingly, other senses of 'kiwi' have 16 tokens which is more than expected. Most of these tokens are either money-related uses, for example again with the New Zealand dollar or referencing a lottery ticket with 'kiwi ticket'. This study is not focused on the other senses due to time constraints and will therefore not go into further detail, but this result was surprising enough to be worth mentioning.

The Wellington data for the written singular maintains the top presence of the New Zealander sense with 32 tokens. As shown in Table 3, the bird sense does have three tokens in the data, but it is still a significant minority. The other senses have even more of a presence with 23 tokens in the written data.

In total, the Wellington Corpus has 74 tokens of the New Zealander type which is significantly more than the three tokens for the bird sense. It also means that the New Zealander sense takes up for roughly 61,7% of the tokens. This difference can be seen clearly in Diagram 3.
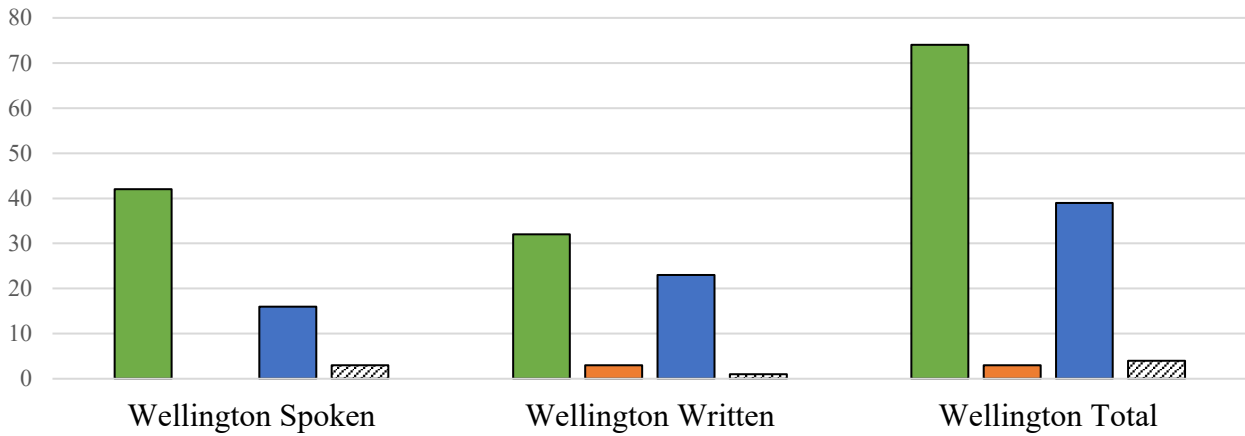
Diagram 3. 'Kiwi' singular in Wellington

The plural data in Wellington is along the same lines as the singular data, just with a lower total token count of 68. The biggest difference from the singular data other than the significantly lower total token count is the lack of ambiguous results and the minor presence of other senses for 'kiwi'. Table 4 shows us that the New Zealander sense has 39 tokens and 22 tokens in spoken and written data, respectively. As with singular data, the bird sense does not appear in the spoken data at all but makes a few appearances in the written data with four tokens.

	Wellington Spoken	Wellington Written	Wellington Total	
Kiwis = New Zealanders	39	22	61	
Kiwis = a type of bird	0	4	4	Total token count:
Kiwis = other sense	3	0	3	68
Ambiguous result	0	0	0	

Table 4. 'Kiwis' plural in Wellington

The significant difference in the combined tokens for written and spoken data does not require visualisation, as the differences are clear. The New Zealander sense comprises

approximately 89,7% of all plural tokens which is an uncommonly large percentage. If we

combine the singular and plural data, we get the results shown in Table 5 and Diagram 4.

	Wellington Spoken	Wellington Written	Wellington Total	
Kiwi/-s = New Zealander	81	54	135	
Kiwi/-s = a type of bird	0	7	7	Total token count:
Kiwi/-s = other sense	19	23	42	188
Ambiguous result	3	1	4	

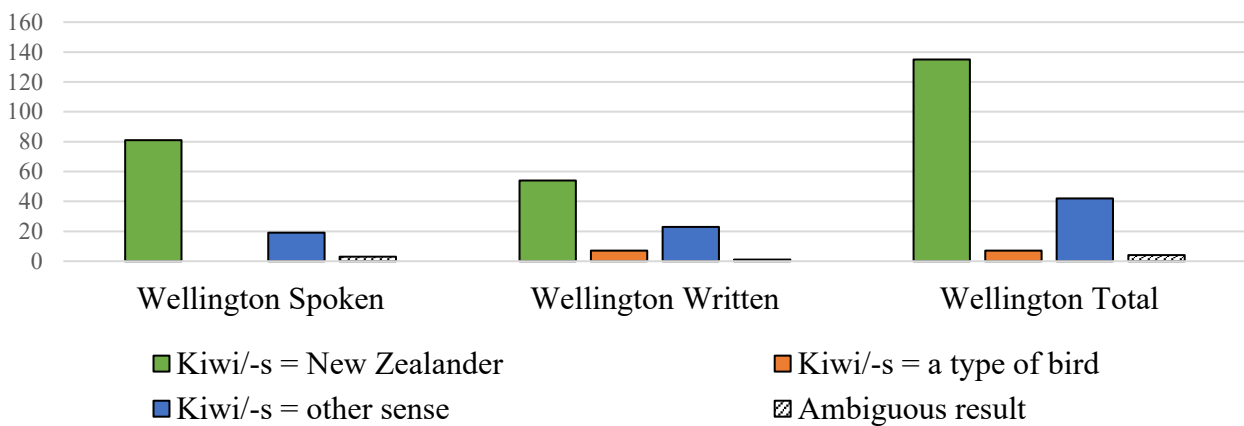*Table 5. 'Kiwi' singular and plural combined in Wellington*

Diagram 4. 'Kiwi' singular and plural combined in Wellington

We can clearly see how the New Zealander sense is the most significant one present

in the Wellington data, with 135 tokens out of 188 resulting in 71,8%, the sense dominates the

other ones. The Wellington data includes a small number of tokens (7) for the bird sense which

even the other senses render minimal with 42 tokens.

3.3. ICE-NZ and Wellington data combined

To further examine the data, both sets are combined into one large set. The total number of

tokens found in both data sets combined is 413. This means that the frequency of *kiwi* and *kiwis*

combined in this set of corpus data is 90.27 words per million. This is around the same size as

Onysko and Calude noted the pair to have in their WSC data, with their total frequency being 96

(2014, 151). However, Onysko and Calude's later data from NZEPC have a frequency of

282.46 (2013, 159). There appears to be a tendency for kiwi to appear more frequently in newspaper texts compared to other written texts and spoken language. Table 6 below showcases the token data from both corpora combined.

	ICE/Wellington Spoken	ICE/Wellington Written	ICE/Wellington Total		
Kiwi/-s = New Zealander	157	72	229		
Kiwi/-s = a type of bird	0	111	111	Total token count:	
Kiwi/-s = other sense	30	30	60	413	
Ambiguous result	10	3	13		

Table 6. 'Kiwi' in ICE-NZ and Wellington data combined

As established by earlier tables, the New Zealander sense is the most prominent one, totalling in 229 tokens out of the 413. The bird sense would seem to be second most common with 111 tokens. However, as was noted in the ICE-NZ analysis, a significant number of tokens come from one source, which distorts the data to an extent. As a result, the number of tokens would be lower if the one source were completely disregarded. The other senses maintain 60 tokens for the combined data which is higher than expected. And finally, the ambiguous hits comprise only 13 of all tokens, most of which come expectedly from spoken data. Diagram 5 serves as a visualisation of the combined data from ICE-NZ and Wellington.
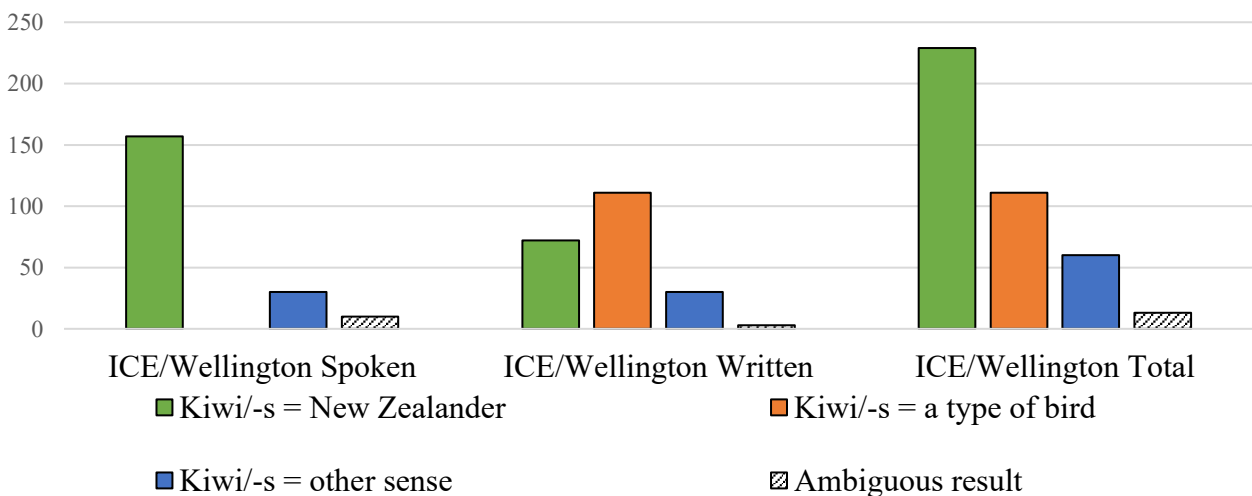

Diagram 5. 'Kiwi' in ICE-NZ and Wellington data combined

4. Discussion

While the results are based on a relatively small sample, they still reveal the changed meaning of *kiwi*. The original sense of the bird seems to have nearly diminished entirely, save for specific contexts while at the same time the newer meaning of a New Zealander sees substantially more use, at least in the data. While this observation is by no means invalid, it should be noted that birds may not be part of the conversation to a large extent and other contexts, such as sports where *Kiwi* is also used, exceed any instances of talk about birds. More data is required to confirm this, but it is speculated that for example possible spoken data gathered may not prove useful as birds are unlikely to appear in day-to-day conversation whereas more written data may more likely include the bird sense of *kiwi*.

The bird sense has not disappeared from the language while its presence has indeed diminished. The trend in NZE is that many fauna and flora are generally not associated with their English names and the Maori counterparts reign supreme. *Kiwi* in its bird sense differs from the New Zealander sense with the fact that *kiwi* is the word used for the species even outside of New Zealand and furthermore even outside of English as is with the Finnish counterpart *kiivi*. In other words, it is impossible without a large revolution within the language that *kiwi* would disappear from the language as there is no replacement. Taking this into consideration it may actually be incorrect to say that *kiwi* has changed meanings; both main uses are still present so perhaps it would be wise to describe this change as an addition to the word instead of a grown preference to the other one because the words appear in greatly different contexts and do not replace each other.

A significant number of the hits for the New Zealander sense were related to sports, which is in line with Onysko and Calude's observation of the same phenomenon (2014, 164-165). This observation is also supported by the OED, as one of the differentiated senses in entry two is a New Zealand sportsman. This does not mean that this is the only context in which *Kiwi*

is used. That said, the significant proportion that is sports-related leads to the conclusion that it is common to use *Kiwi* in sports contexts over *New Zealander*, the latter of which is used by most other countries. It seems that New Zealand English is different in this regard. It would be interesting to see whether foreign sports commentators or journalists use *Kiwi* to refer to athletes from New Zealand, which is something that could be investigated in future studies.

The OED defined *Kiwi* not only as referring to especially New Zealander athletes, but also military personnel. Yet the data does not include a single instance of this type of use. The definition may be correct nonetheless, but *Kiwi* might also not be used to the same extent with soldiers now as it may have been used before. New Zealand military could be something that does not get the same attention in discourse as sports which is highly likely due to 1) New Zealand's remote location and absence of military conflict leading to "irrelevance" of military topics or 2) the more prevalent status of sports contexts in public eye. One other possibility is that this is in fact the change in meaning regarding *Kiwi*; back when New Zealand was growing, and the dialect was forming sports was not necessarily something to think about whereas soldiers and military were more relevant. As the nation grew sports may have come to replace the military use to an extent, which leads to the situation in present day. Regardless, *Kiwi* as New Zealander is the majority use be it sports or military contexts or just generally referencing the people.

This study does not exhaust this topic, as several issues not focused on in this thesis can be investigated. For example, the other senses, such as the kiwifruit (formerly known as Chinese gooseberry) or the kiwi dollar, which were not discussed provide a larger proportion of *kiwi* tokens than the tokenless bird sense in spoken contexts. They also are present to a lesser extent in written contexts, but enough to prompt possible further study. The other senses are widespread in topics as seen from the given examples and it would be interesting to determine their origins. It must be noted that as mentioned in the methodology section not all the "other

sense" tokens are nouns but instead attributives. These could also be interesting to examine more closely in future studies. The problem with these additions to the study would be that the scope would quickly grow and there is the danger of losing focus.

The ambiguous tokens found could also be reviewed alongside other researchers to provide alternative insights into the data, as often ambiguity can be subjective and is one flaw of this thesis. This may however be problematic as the ambiguity may lead to differing views on the context of tokens which would not necessarily be helpful to the study but on the other hand it might give other insights a single researcher is unable to attain by themselves.

More focus could also be placed on the singular and plural forms of *kiwi,* as the use of plural forms of Maori words is a constant discussion within study of NZE. In the case of *kiwi*, even though it is an exception to the general rule, there may be differences in which contexts the plural form with *-s* is used and where not, but this was not the focus of this study. The data also shows a slight preference to using the singular form in both ICE-NZ and Wellington, not regarding the data anomaly with the bird sense in ICE-NZ. This could be investigated further with more data to determine if the singular is truly preferred with plural being just a possibility. There is also potential for further study with how the plural *kiwis* became an exception to the rule about non-plurality of Maori words.

The data of the study is adequate for the scope, but some inclusions could be made. For example, a separate data set for newspaper data could be studied in the light of this study. As mentioned before, the ONZE data would greatly increase the relevance and precision of this thesis. This is the case because ONZE is the only source that maps the whole progress NZE has made as a language from the beginning to the end. The distortion of one source text in the ICE-NZ data is something of a problem for the data. While it contributes to the data, it manufactures a significant portion of the bird sense words in the study. This leads to the questioning of the reliability of ICE-NZ while not completely rejecting the use of the corpus. The shortcomings of

ICE-NZ are not something that discredits this thesis as Wellington compensates with providing unproblematic data. The problems of ICE-NZ are also fully considered and therefore do not cause major issues even if the data distortion is visible in the charts; if this problem is known, the unusual charts can be explained.

One excellent addition the to the data alongside ONZE would be any corpora that are more contemporary than ICE-NZ and Wellington. Both corpora used in this study are strictly consisting of modern data compared to ONZE, but the data is from the late 1980s and the 1990s which means the oldest data in the corpora is already 30 years old. I highly doubt there has been any massive change in the 2000s, 2010s and now the 2020s but having newer data closer to our time would help with relevance and would broaden the study. One example of a possible change, and this is completely hypothetical, is that perhaps military topics have gained relevance in the 30 years and the soldier context has grown in commonness. Another completely hypothetical change would be that maybe there has been an event or disaster relating to the kiwi birds that has made the sense more prevalent in newspapers. These hypothetical changes may or may not be accurate, but without newer data it is impossible to know. That said, it is likely that the results of this study hold up with contemporary data as to my knowledge, nothing revolutionary has happened.

As this thesis consists solely of corpus study, it should be noted that fieldwork that could be done might affect the results. By travelling to New Zealand, one could gather information from face-to-face conversations, sports broadcasts, or contemporary newspapers to list some possibilities which are not limited to these. Of course, this would require significantly more work compared to a corpus study, which would also benefit from travelling as mentioned before, and as this is one of the strengths of using corpora. Corpora are by no means perfect for this kind of study (as is evident from ICE-NZ) as you are working with a locked set or sets of data unless you compose your own corpus which would definitely require either fieldwork or

compiling a significant amount of data from written/online sources. Of course, corpora can be updated so "locked" is a bit misleading. However, as the data is generally unchangeable at the time of a single study, this may result in some phenomena falling outside of the corpora data. Then again, this is true for every kind of study, as no method is omnipotent. By broadening the study and using different methods for a future iteration these problems could be marginalised and the study made more accurate. Even with the limited resources available, this thesis presents relevant results.

5. Conclusion

The original intent of this thesis on the word *kiwi* in New Zealand English was to examine the historical progress regarding the change in meaning which proved to be very difficult with the tools available. So, the new question asked in this thesis regards the significance of the newer sense. Results show a clear difference in the commonness of the two senses of *kiwi*, with the New Zealander sense being the more prominent one by a large margin. This leads to the answer for the second part of the question on the sensibility of future research on the history of *kiwi*. The New Zealander sense of *kiwi* is also very prevalent in sports contexts according to the data, which is something to consider in future research, while other senses not investigated in this study also contribute to the use of *kiwi* and are worthy of a study as they are combined as they are about as common in written contexts and more common in spoken contexts than the bird sense.

The observations made here would indicate that there has not been a change in the meaning of the word but instead a shift in relevance as both senses are still very much used in addition to other senses of the word. This should provide a good starting point for further study. For example, the inclusion of more newspaper data would be a good addition and the mentioned ONZE data would allow the examining of the historical development. One could also, similarly

with Onysko and Calude, study different Maori words in addition to *kiwi* with the ONZE data

and examine how unique this change with *kiwi* is within New Zealand English. Overall, access to

ONZE data would be the most beneficial continuation for this study due to the data including the

entire evolution of NZE, allowing a more complete understanding of the use of *kiwi* throughout

history. As discussed before, alternative research methods could also provide a wider

understanding of the situations with *kiwi*; field work could prove very useful in getting spoken

data. The other uses of *kiwi* could also be investigated thoroughly in further research which may

broaden the understanding of the current use of the word. While this study contributes to the

field with a very specific topic that same topic could be broadened in several ways which makes

approaching further studies from multiple different angles possible.

References

English corpora resources. 2019. [Internet] Victoria University of Wellington. Available from

https://www.victoria.ac.nz/lals/resources/corpora-default [Accessed 24 November

2019]

Hay, Jennifer. Maclagan, Margaret. Gordon, Elizabeth. *New Zealand English*. 2008. Edinburgh

University Press Ltd. 22 George Square, Edinburgh.

"kiwi, n." OED Online. September 2019. [Internet] Oxford University Press, Available from

www.oed.com/view/Entry/103797. [Accessed 24 November 2019]

Onysko, Alexander & Claude, Andreea. "Comparing the usage of Māori loans in spoken and

written New Zealand English: A case study of *Maori*, *Pakeha*, and *Kiwi*". *New*

Perspectives on Lexical Borrowing. 143-169. 2014. Walter de Gruyter, Inc.,

Boston/Berlin.

Trudgill, Peter. 2004. *New-dialect formation*. Edinburgh University Press Ltd. 22 George Square,

Edinburgh.