

Gunnar Holmberg  
[g.holmberg@live.se](mailto:g.holmberg@live.se)

BACHELOR THESIS  
(JAPK11) Spring semester 2010

# Japanese, Austronesian and Altaic

## A study of possible connections



LUND UNIVERSITY

Centre for Language and Literature, Japanese Studies  
Supervisor: Lars Larm

# **Abstract**

The question of the origin of Japanese is a mystery still unsolved, proven by the academic world teeming with various contradictory hypotheses. This thesis evaluates whether the Austronesian connection with Japanese is plausible in the face of the by amount of evidence much more firmly established Altaic-Japanese hypothesis, showing that measured by linguistic evidence there may be a connection to both.

Through analysing lexical items and their semantic domains, as well as historical and archaeological facts, this thesis attempts to determine whether the two hypotheses are completely incompatible or not.

**Keywords:** Austronesian, Altaic, Japanese, Genetic relationship

# Conventions and Abbreviations

## Romanization

In this thesis, the Revised Hepburn system is used for names of authors, places and historical periods, while *kunreishiki* will be used in transliteration of Japanese expressions. The Hepburn system transliteration resembles English pronunciation, and is the most popular system in popular use today. *Kunreishiki* is popular among linguists, and the sources of vocabulary which this thesis builds upon apply this system. The main difference between the two are the syllables *tsu, shi, chi, fu* and *ji* in the Hepburn system are written *tu, si, hu, zi/di* in *kunreishiki*.

Words written in CAPITALS signify semantic meaning.

*Italics* are used for the titles of books as well as Japanese and other non-English words in the text.

‘Single quotes’ represent the meaning of a word as translated by the source.

## Abbreviations

K	Modern Korean
Jp	Modern Japanese
OJ	Old Japanese

## Table of contents

Abstract.....	1
Conventions and Abbreviations.....	2
1. Introduction.....	1
2. Approach.....	1
3. History.....	2
3.1. Jōmon (ca 13000–250 BC).....	3
3.2. Yayoi (ca 250 BC–250 AD).....	3
4. The question – an overview.....	4
5. The Altaic hypothesis.....	5
5.1. Arguments against the Altaic hypothesis.....	6
6. The Austronesian hypothesis.....	7
6.1. Arguments against the Austronesian connection.....	9
6.2. Austronesian vs. Altaic – Robbeets's evidence against Benedict's.....	10
7. Basic vocabulary.....	11
7.1. Robbeets and the Swadesh list.....	11
7.2. Benedict and the Swadesh list.....	12
7.3. Cognates in specific semantic domains.....	13
8. Conclusion.....	14
9. Future research.....	15
Bibliography.....	16

## 1. Introduction

What happened in the Japanese islands millennia ago may seem of little importance, but if history has proven anything it is that the origin of a people can play an essential role in their view of themselves and their relation to others.

It has been called *the champion of hypotheses*: the Japanese language has remained the only major language of which the origin is still a mystery. There is no other language in the world with which so many attempts at determining ancestry have been made, few of which have proven fruitful.

Why is it that such a large language has such a vague history compared to that of, say, most European languages? First of all, if Japanese is related to Korean, which many believe, the split of the two languages must have happened millennia ago – a time from which no written records remain. Compared with the Romance languages' split inferred to about 1500 years ago, a time from which several written records remain, the difference becomes evident.

Furthermore, a vast amount of Chinese loan words have eliminated their Japanese counterparts, further obscuring the matter. It is possible that the language, by time and influence of other languages, has changed so much from its ancestor that the origin may never be revealed.

Theories on the question are often presented with great certainty by their author, and hypotheses are often claimed to be proven true (or false) once and for all. However, there is to this date no general consensus among scholars<sup>1</sup>.

This thesis compares a leading hypothesis of the possible Altaic connection to Japanese with a slightly less accepted theory of an Austronesian connection in an attempt to find out whether the Austronesian hypothesis is truly – as claimed by many – implausible or if it has simply suffered inadequate representation.

## 2. Approach

In this thesis, a comparison of the evidence for and against the Altaic-Japanese as well as the Austronesian-Japanese hypotheses is conducted, evaluating arguments such as how cognates correspond to basic vocabulary, as well as historical facts.

Non-linguistic evidence is used in order to explain hypotheses and to evaluate other non-linguistic arguments. Note should be taken that historical evidence concerning Japan is a subject where the facts are constantly revised. The year at which an occurrence has been

---

<sup>1</sup> Save the fact that Ryukyuan is a dialect of Japanese, which is today generally agreed upon.

inferred should be taken *cum grano salis* as it is only an approximate.

A comparison of two theories, both of which are renowned among scholars in the area is used as base when determining whether one may simply be showing similarities with Japanese due to superstratum interference over a long time.

Altaic and Austronesian hypotheses were compared, mainly with Paul K. Benedict's *Japanese/Austro-Tai*<sup>2</sup> and Martine Irma Robbeets's *Is Japanese related to Korean, Tungusic, Mongolic and Turkic?* as basis for the comparison. Benedict's book, upon its release, sparked new life into the hypothesis of a Japanese-Austronesian connection, proposing an amount of cognates that by sheer size is hard to refute as mere coincidence. Robbeets's work is currently leading among proponents of the Altaic connection, boasting no less than 2055 thoroughly evaluated proposed cognates.

The fact that many linguists have proposed different combinations of the two theories was a further reason for the choice.

This thesis compares the lexical items proposed as main evidence in each hypothesis, with focus on semantic domains, in an attempt to see what the connection (if any) Austronesian may have with Japanese and Altaic.

### 3. History

While archaeological findings, historical facts, genetic surveys etc. can certainly play an integral part in determining distant genetic relationship, they can also be deceiving, and must never be considered core evidence of a hypothesis (Campbell 2004:356). Linguists tend, however, not to confine themselves solely to linguistic evidence when supporting or denouncing a theory. An overview of the Japanese archipelago's early history, as well as the surrounding lands, is helpful in understanding what problems one will encounter when trying to determine the ancestors of the Japanese language compared to say, languages of Europe.

In the case of Japanese and distant genetic relationship, focus lies mainly on two questions: what language(s) was spoken in the Japanese archipelago during the Jōmon period, and in what way did it change in the Yayoi period? The proof that currently exists is capable of pointing linguists in more than one direction.

---

<sup>2</sup> While Benedict believes that the genealogy of Japanese and other, according to him, Austronesian languages have their roots in Austro-Tai, focus of his research lies on relating Japanese to Austronesian. His theory can therefore be called Austronesian-Japanese as well as Austro-Tai, and will be referred to as such in this thesis.

### **3.1. Jōmon (ca 13000–250 BC)<sup>3</sup>**

Jōmon means straw-rope pattern, as the Jōmon people used straw-ropes to decorate pottery. The pottery in Japan is the oldest known to archaeologists. What today constitutes the Japanese archipelago was a part of the Asian mainland until around 13000 years ago. There are several suggestions as to the origins of the nomadic population living in Japan at the time. Some findings indicate that it was the same people who occupied the areas around China and Mongolia at that time, but people from southern Asia also seem to have journeyed there, obscuring the picture.

During the latter half of Jōmon, the population appears to have gone from nomadic to sedentary while still living mainly as hunter-gatherers. This is rather unusual, as most societies through history have become sedentary as they adopt agricultural techniques. It is believed that the Japanese islands were so abundant in food, which, combined with the early knowledge of pottery allowed the Jōmon people to live relatively stationary without farming. The sedentary societies in Japan do not appear to have had much contact among each other, and one could speculate that this may have led to differing languages throughout the archipelago.

The Jōmon people appear to have had some contact with the Korean peninsula, and may have been aware of agricultural techniques, but the techniques were not very effective at the time, and the Jōmon people probably saw little advantage in adopting them (Diamond 1998).

### **3.2. Yayoi (ca 250 BC–250 AD)**

Around 400 BC a great change took place in Japan, the particulars of which may be an important link in determining where the Japanese people and their language come from.

While the civilizations of Asian mainland had been developing agricultural techniques for millennia, the Jōmon people remained mainly hunter-gatherers up until this point. An explanation may be that in the relatively cold climate of Korea, through which most exchange with the mainland is thought to have taken place, cold-resistant but less effective techniques were being used, and thus were of little or no advantage to the Jōmon people. However, around 400 BC there seems to have been a great change which led the people to adopt irrigated rice paddies among other things. This induced a shift from being a hunter-gathering population into a farming one, as well as a dramatic increase in population<sup>4</sup>.

Little is known by what means these technological traits came into use in Japan, and

---

<sup>3</sup> Dates of Jōmon and Yayoi period differ in different sources. The most recent one available to the author was chosen (Ottosson & Ekholm 2007).

<sup>4</sup> The farming techniques used during the Yayoi period originated in China, though according to archaeologists they were probably mediated to Japan through Korea.

whether this could be explained by an invasion from Korea, or simply an immigration. Perhaps the people in Japan simply established trade on a new level with the Korean peninsula. According to Ottosson & Ekholm (2007) the most probable theory is that several smaller groups immigrated during a long period of time from the Korean peninsula. It is important to note that other ways have been suggested for the rice expansion into Japan, for example through Taiwan – which would support the Austronesian hypothesis (see Barnes, 1999).

## 4. The question – an overview

Shibatani (1990) tries to sum up the state of the question of genetic affiliation to Japanese. Expanding a list originally constructed by Kamei (1961), Shibatani breaks up the different hypotheses into six groups:

1. Hypotheses that connect Japanese with languages of North Asia.
2. Hypotheses regarding a South Asian connection.
3. Hypotheses on connections with Indo European languages.
4. Other hypotheses (eg. connection with Basque and Sumerian.)
5. The hypothesis that believes Japanese to be an Altaic superstratum upon an Austronesian substratum.
6. A hypothesis that Japanese is a mix of Austronesian and Altaic.

The theory of an Altaic superstratum and Austronesian substratum basically means that an Austronesian language was spoken in early Japan, but later on an Altaic language took over. Through learning imperfections, cultural terms etc. traces of the Austronesian language remained, though the Altaic language dominates as a superstratum. The mixed language hypothesis claims that Japanese has more than one single ancestor, and that the languages from which it descends both have great influence upon each other.

Shibatani then goes on to give an account of the more successful theories<sup>5</sup>. Shibatani's work is, however, from 1990, and therefore fails to include one hypothesis which through the research of Christopher Beckwith has given cause to much debate recently: the theory of a Japanese-Koguryo connection. Beckwith claims that Japanese descends from Koguryo, which in turn is connected to old Chinese, and perhaps distantly to Tibeto-Burmese.

Stating that “Upon the scholarly discovery of the Koguryo language in 1907, and ever since,

---

<sup>5</sup> According to Shibatani, groups 3 & 4 are said to be unsuccessful, and will be excluded from further scrutiny in this thesis.

there has never been any doubt [according to Beckwith] that it is genetically related to Japanese" (2006), Beckwith claims that the Koguryo hypothesis currently holds the greatest support, but at a closer look it has received a great deal of criticism from other experts. Criticism mainly consists of the points that reconstruction of Koguryo is based on toponyms, which by many is considered insufficient as data, and uncertainty in whether Koguryo is even a language of its own or just a dialect of old Korean<sup>6</sup>.

## 5. The Altaic hypothesis

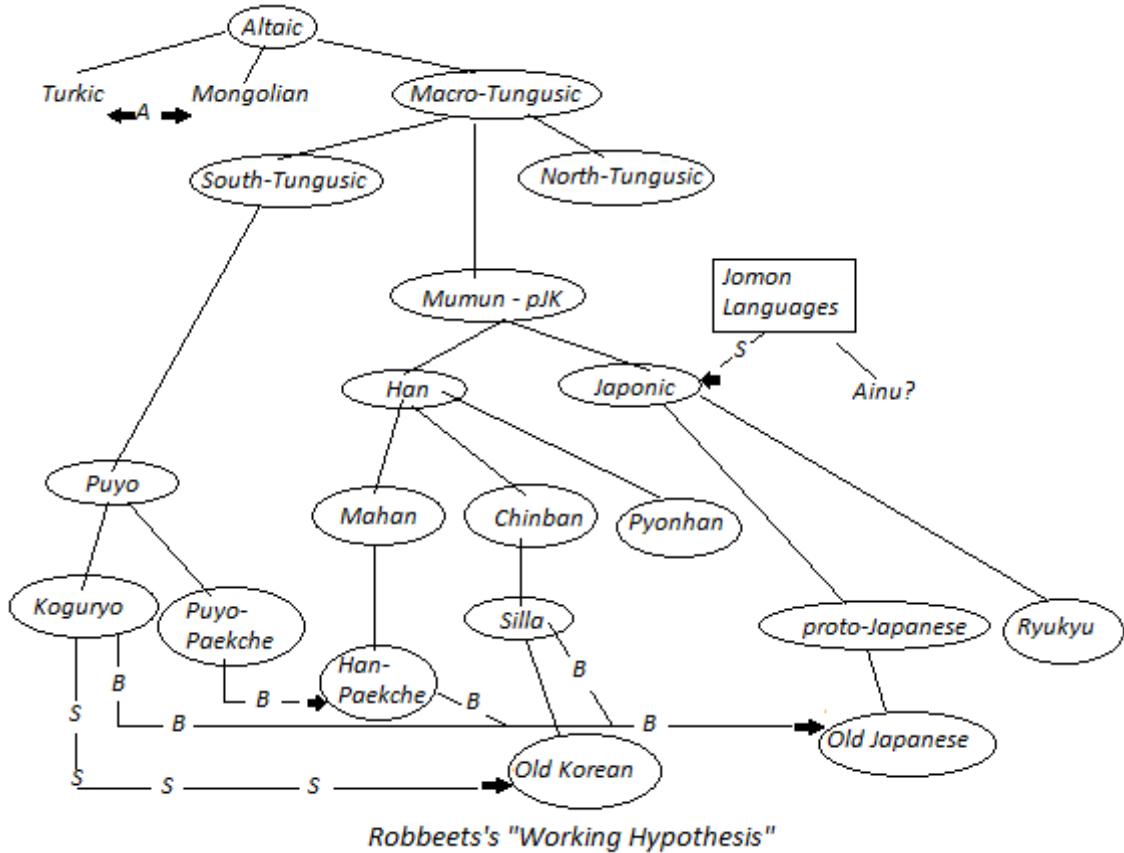
The history of attempts at attesting a Japanese-Altaic relationship stretches farther than a century back in history. In 1832, little less than 200 years ago, von Siebold proposed said connection.

According to the Altaic hypothesis, the language spoken in Japan during Jōmon completely disappeared through the immigration of a higher culture from the Korean peninsula. It is known to linguists that when less advanced civilizations are exposed to a higher culture, there tends to be either superstratum interference, or language shift – neither of which would necessarily refute a Japanese-Altaic connection.

The Altaic-Japanese connection is perhaps the most extensively investigated hypothesis regarding the origin of Japanese. One of the leading attempts at proving this connection is Robbeets's 2005 book, based on a doctoral dissertation in 2003, where data from other scholars' research is assembled and evaluated. Out of 2055 gathered etymologies, "359 etymologies [...] show a perfect phonological fit for the initial consonant, the medial vowel and the medial consonant of the Japanese entry." (2005:378).

---

<sup>6</sup> *Journal of Inner and East Asian Studies*, volume 2-2 (2005)



*B = Superstratum interference (borrowing).*

*S = Substratum interference.*

*A = Adstratum interference, both languages have equal influence upon each other.*

### 5.1. Arguments against the Altaic hypothesis

While Benedict argues for an Austronesian relationship to Japanese where resemblances with Korean are explained by chance or loan, little is said about the possibility of Austronesian being a substratum. On the other hand, an example is given which Benedict considers a loan from Japanese to Korean: Jp. *kuma* ‘bear’, which has an “impeccable Austro-Tai genealogy”. *Kuma* can actually be found in Robbeets’s full list of (unsifted) cognates, but apart from the Korean resemblance, a more tentative answer is given where it is compared to the proto-Tungusic word for ‘seal’, the proto-Turkic word for ‘beaver, otter’ among others.

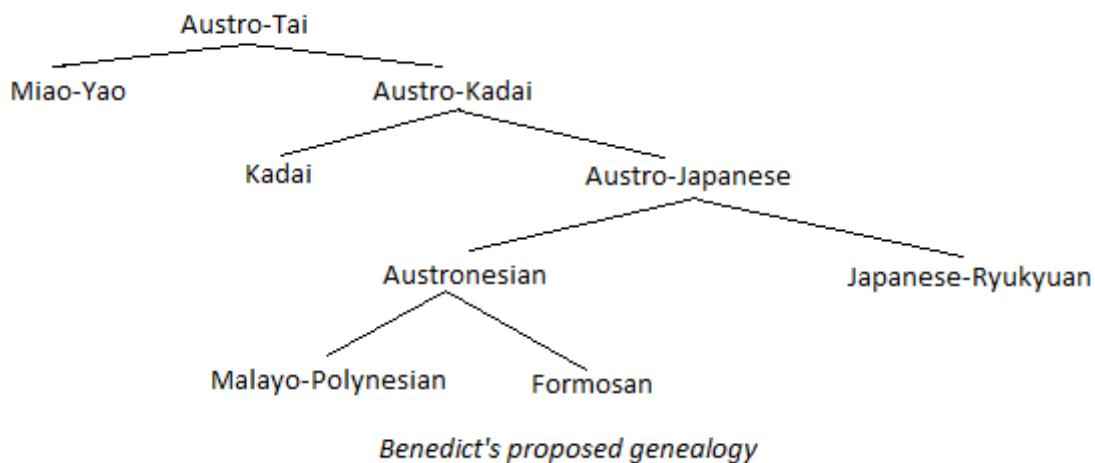
A major obstacle for historical linguists is the possibility of superstratum interference. Regarding the Altaic hypothesis, Robbeets admits that such a risk exists, that “extensive borrowing at a certain period of time can result in strata of loanwords that exhibit systematic

sound correspondences”, but in her defence adds “Maybe these opponents [to the Altaic hypothesis] are right, but I am convinced that arguing for or against genetic language relationship after individual examination of every single entry suspect of being a loanword will prove to be a much more profitable approach than abandoning the entire enterprise *a priori* on grounds of ‘wholesale borrowing’.” (2005:188).

## 6. The Austronesian hypothesis

In contrast to the Altaic-Japanese hypothesis, an Austronesian connection with Japanese was first suggested in 1924 by a linguist named Labberton, about a century after von Siebold’s theory on Japanese-Altaic. Note that on top of there being a longer history of Altaic-Japanese research among western scholars in total, it has also had proponents in larger number than the Austronesian hypothesis historically has.

The Austronesian hypothesis basically states that Japanese descends from languages with origin in southeast Asia, and if one is to follow Benedict’s view it then travelled to the island Kyushu in southern Japan by way of Taiwan.



Miao-Yao languages are nowadays called Hmong-Mien.

While there aren’t that many who believe in a “pure” Japanese-Austronesian connection, variations of the hypothesis such as the Altaic superstratum on top of an Austronesian substratum have a noteworthy amount of proponents.

In a 1992 review of Benedict’s “Japanese/Austro-Tai”, David Solnit makes an interesting observation upon comparing lexical items from a survey of Altaic with those of Benedict’s. Solnit compares 89 lexical items from the Altaic hypothesis with those of Benedict’s “whose gloss includes the meaning in question, even if the Japanese word does not exemplify the

meaning in question”. The comparison yielded the following results:

Austronesian cognates only	31	28%
Altaic cognates only	45	40%
competing <sup>7</sup>	25	23%
missing altogether	10	9%
total	111	100%

Only comparing the data at “face value”, without assessing reliability of the comparisons, Solnit’s survey alone is not enough to draw any conclusions from. While not a solid proof, Solnit states it could be an indication that there has been contact of some sort or the other with both Austronesian and Altaic languages in the past. A similar comparison done in this thesis yields similar results (see 7).

---

<sup>7</sup> “In the tables, ‘competing’ means that a single Japanese item is compared to Altaic by Starostin and to Austronesian by Benedict.” (Solnit 1992). Furthermore, the reason that the numbers of compared items exceed 89 is that there are at times more than one proposed etymon for one semantic meaning.

## **6.1. Arguments against the Austronesian connection**

Benedict's book on the matter does not apply as strict criteria as Robbeets's does for Altaic cognates. As his hypothesis is meant to show through preliminary evidence that Austronesian and Japanese are related, "Other apparent correspondences, each involving a 'problem' of sorts, are included in the text for heuristic purposes; it is anticipated that a good many of these, perhaps even a majority, will in time be authenticated."

(1990:18).

Robbeets claims that while proponents of Austronesian theories hitherto have failed to produce more than 500 likely cognates, her thesis presents over 2000 "Japanese entries along with various mythologies relating the entry to Korean or Altaic" (though only 359 live up to her stricter criteria). Also, "This statistical fact alone makes the Altaic hypothesis a more promising theory to test than the Austronesian hypothesis for Japanese". A valid argument indeed, but there are possible reasons for this that don't necessarily prove the Austronesian connection invalid. It may be that the Austronesian hypothesis is simply underrepresented among linguists, as claimed by Benedict (1990:1-2). One could also compare Japanese to Chinese and find regular sound correspondences and likely cognates (though this, unlike Altaic hypotheses, is known for certain to be due to mass borrowing and not genetic relationship). Unfortunately, to the author's knowledge, apart from substratum, hybrid or mixed language hypotheses, there have been no subsequent attempts at connecting Japanese and Austronesian since Benedict's in 1990. Robbeets too, fails to refer any attempts made after Benedict's that suggest genetic affiliation with Japanese – other than the highly controversial mixed language hypotheses<sup>8</sup>.

Another of Robbeets's arguments is that the Austronesian connection supposedly is incompatible with historical facts. If one is to follow Lyle Campbell's view of how to properly determine distant genetic relationships, then "Shared cultural traits, mythology, folklore, technologies and gene pools must be eliminated from arguments for linguistic relationship." (2004:356). There appear to be few linguists (if any) who completely exclude such facts (especially when they support their theories), and while they shouldn't be considered peremptory, they do indeed add to the believability of the hypothesis.

Another argument against an Austronesian origin is that it bears very little resemblance to Ainu:

"Biological and archeological findings suggest that Ainu may be derived from a

---

<sup>8</sup> Robbeets, (2005:36–8).

Jōmon language. If the Jōmon languages indeed were Austronesian, as suggested in the Austronesian hypothesis, then Ainu, a descendant of a Jōmon language is probably a sister language of Austronesian. If Japanese consists of Austronesian elements or if it has an Austronesian substratum, it is reasonable to expect that Austronesian is closer to Ainu than [sic] to Japanese.” (Robbeets 2005:37)

This is mainly a non-linguistic argument, and can be responded to in the same manner: the people of the Japanese archipelago were not united, and were also sedentary possibly as early as 5000 BC. The various people residing in the Japanese islands may in fact not have had contact with each other, and if a foreign language was indeed introduced early on in one part of the island it does not necessarily mean that the whole island would adapt it. Also, the introduction of higher culture and new technology is not always immediately incorporated by the indigenous population. Archaeological findings indicate that the spread of Yayoi culture didn't fully reach northern Japan at first. Little is known of how the Yayoi people interacted with the Jōmon people, but there are remnants of Jōmon culture even in places where the new technologies were adopted, suggesting a gradual merging of the two.

### ***6.2. Austronesian vs. Altaic – Robbeets's evidence against Benedict's***

“Wholesale borrowing” cannot be excluded as a problem with Benedict’s theory either. For example, prolonged contact after the language split is proposed as reason for there being many more attested cognates between Japanese and Formosan than the other branches. What isn’t considered, however, is that such long-term contact can also induce borrowing – whether there is genetic relationship or not.

One major difference for which one might criticize this comparison is that in lack of previously established evidence to build his theory upon, Benedict had a rather broadened scope in the search for cognates. The Altaic hypothesis, on the other hand, builds on a long history of careful scrutiny, and Robbeets boasts many times more proposed cognates than Benedict. However, Robbeets’s “core evidence” with its 359 proposed etymologies aren’t that far from Benedict’s full list of 256. This survey will consider the full set of etymologies proposed by Benedict with the core evidence of Robbeets’s, as no such narrowing down of proposed evidence was done by Benedict.

## **7. Basic vocabulary**

When considering genetic relationship of languages, the existence of basic vocabulary is considered a strong argument. Basic vocabulary is supposed to consist of culture-free words, body parts and other words that are thought to be needed for survival in any civilization.

Introduced in 1955 by Morris Swadesh, the 100-word Swadesh list is still the most widely used.

The Swadesh list has not gone without criticism. For example, while the peoples of Central America had dogs before meeting the Europeans, the dogs from Europe differed greatly in appearance. Consequently, the European word for dog was often borrowed, eventually leading to the original word disappearing.

In spite of such criticism, this thesis will make a comparison based on the Swadesh list, due to it still being the more widely used list of basic vocabulary, and that Robbeets uses it as part of her main evidence.

### 7.1. *Robbeets and the Swadesh list*

Robbeets considers basic vocabulary cognates to be a strong evidence, the following list shows what words on the Swadesh 100-word list have a Japanese-Altaic cognate.

1. **I** 2. **you** 3. we 4. this 5. that 6. what 7. who **8. not** 9. **all** 10. **many** 11. **one** 12. **two** 13. **big** 14. long 15. **small** 16. woman 17. **man** 18. **person** 19. fish 20. bird 21. dog 22. louse 23. **tree** 24. seed 25. leaf 26. root 27. **bark** 28. **skin** 29. flesh 30. **blood** 31. **bone** 32. egg 33. grease 34. horn 35. tail 36. feather 37. hair 38. head 39. ear 40. eye **41. nose** 42. **mouth** 43. **tooth** 44. tongue 45. claw 46. **foot** 47. knee 48. **hand** 49. **belly** 50. neck 51. breast 52. **heart** 53. liver 54. drink 55. **eat** 56. **bite** 57. see 58. hear 59. know 60. **sleep** 61. die 62. **kill** 63. **swim** 64. fly 65. **walk** 66. **come** 67. lie 68. sit 69. stand 70. **give** 71. **say** 72. **sun** 73. moon 74. **star** 75. water 76. rain 77. stone 78. sand 79. **earth** 80. cloud 81. smoke 82. **fire** 83. **ash** 84. **burn** 85. path 86. mountain 87. **red** 88. green 89. **yellow** 90. **white** 91. black 92. night 93. **hot** 94. cold 95. say 96. good 97. new 98. **round** 99. dry 100. name<sup>9</sup>

The 45 words marked in bold have a correspondence in the core evidence lexicon. It must be noted that these are only the cognates among Robbeets's core evidence of 359 etymon that are on the Swadesh list. When comparing the full evidence of 2055 lexical entries, 92 of the Swadesh list words are found according to Robbeets.

<sup>9</sup> Robbeets did not list '46. foot' and '48. hand' as core evidence that fit the Swadesh list , yet *ashi* (foot, leg) and *te* (hand) can be found with Korean etymology in the complete list of core evidence vocabulary that follows (2005:399, 402) and is therefore included in this comparison.

## 7.2. Benedict and the Swadesh list

Unlike Robbeets, Benedict does not make any argument based on representation in the Swadesh list. The author of this thesis has therefore taken the liberty of doing so, comparing the full 236 lexical entries which semantically correspond to the Swadesh 100-word list:

1. I 2. you 3. we 4. this 5. that 6. what 7. **who** 8. not 9. all 10. many 11. **one** 12. **two** 13. big 14. long 15. **small** 16. **woman** 17. man 18. person 19. **fish** 20. **bird** 21. **dog** 22. louse 23. **tree** 24. **seed** 25. **leaf** 26. **root** 27. bark 28. **skin** 29. flesh 30. **blood** 31. **bone** 32. egg 33. grease 34. **horn** 35. **tail** 36. **feather** 37. **hair** 38. head 39. ear 40. **eye** 41. nose 42. **mouth** 43. **tooth** 44. tongue 45. claw 46. **foot** 47. knee 48. **hand** 49. **belly** 50. neck 51. **breast** 52. heart 53. liver 54. **drink** 55. **eat** 56. **bite** 57. see 58. hear 59. know 60. sleep 61. **die** 62. kill 63. swim 64. **fly** 65. walk 66. come 67. lie 68. sit 69. stand 70. give 71. say 72. **sun** 73. moon 74. **star** 75. **water** 76. rain 77. stone 78. **sand** 79. **earth** 80. cloud 81. smoke 82. **fire** 83. ash 84. burn 85. path 86. **mountain** 87. red 88. **green** 89. **yellow** 90. white 91. black 92. **night** 93. hot 94. **cold** 95. say 96. good 97. new 98. **round** 99. dry 100. **name**

Out of these 45 items, 20 fit under the same semantic meaning as the Altaic proposals<sup>10</sup>. In several cases for both lists, more than one suggested item fits semantically with the Swadesh list; for example OJ *ki* ‘fang/tusk’ and Jp. *ha* ‘tooth’ are both listed semantically as “TOOTH”.

When looking only at the proposed Japanese word for respective Swadesh list word, 17 items have the same Japanese suggestions as claimed cognates for Altaic. As there are often more than one suggested cognate for each of Swadesh’s words, one could choose to look at the cases where the same Japanese word is *not* presented, and end up with 9 items out of 20. There are several items that can be considered applying to the Swadesh list of basic vocabulary that are not among the Altaic cognates.

All in all, these results make a strong point in favour of the Austronesian hypothesis being plausible. While the criteria are more relaxed than with the Altaic hypothesis, further analysis of Austronesian languages may reveal cognates that can compete under equal circumstances.

## 7.3. Cognates in specific semantic domains

Claiming it a strong evidence for Austro-Tai, in the conclusion of his thesis Benedict presents nine cognates in agricultural terminology:

---

<sup>10</sup> Conflicting with Altaic: 1. I 11. one 12. two 15. small 23. tree 28. skin 30. blood 31. bone 42. mouth 43. tooth  
46. foot 48. hand 49. belly 55. eat 56. bite 72. sun 74. star 79. earth 82. fire 89. yellow 98. round

‘paddy field’: Jp. *ta* (EARTH)  
‘to plant’: Jp. *ue-* (PLANT)  
‘rice plant’: Jp. (*s-*)*ine* (RICE)  
‘unhulled rice’: Jp. *yone* (RICE)  
‘rice’: Jp. *kome* (RICE)  
‘hulled rice’: Jp. *momi* (RICE)  
‘millet’: Jp. *awa* (SWAMP)  
‘mortar’: Jp. *usu* (MORTAR)  
‘winnower’ Jp. *mi* (WINNOW)

*Words in capital letters represent semantic domain, as assigned by Benedict.*

Out of these nine words only one, ‘to plant’: Jp. *ue-*, can be found among the 359 core vocabulary glosses of Robbeets’s (as Jp. *ueru*, 2005:394). When compared with her full list of 2055 glosses, there are 5 more competing words: ‘mortar’, ‘paddy field’ (as ‘ricefield’), ‘rice plant’ (as Jp. *ine*) and ‘rice’. Words that would fit under the semantic domain of agriculture are remarkably scarce in the core evidence of the Altaic hypothesis. This could be an indication that the culture which brought farming techniques to the Japanese islands was in fact not that of the Korean peninsula.

Benedict states many other specific cognates that he believes promote the idea that the language of the higher culture immigrating to Japan was of Austronesian origin. The rest of the evidence is, however, spread thin over different semantic domains, and Benedict admits that focus lies on agricultural terminology.

If the cultivation of rice spread from southern China to Korea, and to Japan by way of Korea, then the resemblances with south-east Asian words in Japanese ought to have arrived the same way. This raises the question as to why such correspondences are absent in the Altaic hypothesis, in other words not corresponding to Korean which it claims is the linguistically closest relative to Japanese.

## 8. Conclusion

When compared to Robbeets’s core evidence, 25 words on the 100-word Swadesh list were found to have a corresponding lexical item only among attested Austronesian-Japanese cognates. All in all, the Austronesian hypothesis presents more cognates than can easily be

rejected as mere chance. Also, the example of Jp. *kuma* ‘bear’, appears to have a better established Austronesian genealogy than the Altaic suggestion, and points towards influences from Austronesian languages in Japanese.

The Japanese-Altaic theory arguably has the strongest support among scholars today, and the evidence to back it being so. That being said, possible contact between Austronesian languages and Japanese do not seem to have been explored to the extent that one would hope for in proving it invalid. This thesis aims to show that there are elements of the Austronesian hypothesis that aren’t conclusively disproved, and that the hypothesis presents evidence which challenges the Altaic hypothesis in some areas. Something which should be specifically looked into is the quality of the evidence which connects Japanese and Austronesian terminology for agriculture, as well as the manner of contact between proto-Japanese and Formosan languages, to determine whether resemblances are due to borrowing, chance, or genetic relationship.

Based on the information gathered in this thesis, the deduction can be made that there today isn’t enough evidence to confidently attest genetic relation between Japanese and Austronesian, but that there are many indications of an Austronesian substratum under an Altaic superstratum (the Altaic hypothesis still presenting a stronger case altogether).

Evaluations should be made on whether the evidence in favour of the Austronesian connection holds its ground even with stricter criteria and research conducted in search of cognates.

Many of the arguments against an Austronesian-Japanese connection are based on non-linguistic evidence, evidence which in reality tells next to nothing about what language was actually spoken at the time, and can often be retorted in the same manner. It is the author’s belief that if one is to validate or disprove a hypothesis of genetic relation between languages, it ought to be by means of linguistic evidence.

While interdisciplinary research such as archaeology and genetics may play an important role in the question of where the Japanese language comes from, linguistics will most certainly have the last word.

## 9. Future research

Further investigation of possible cognates with Austronesian, as well as in-depth evaluation of the currently proposed evidence would be the obvious way to continue the research conducted in this thesis.

Morphological evidence as well as vowel systems are not dealt with in this thesis, and could

prove invaluable to the future of the question.

Furthermore, recent advances in linguistic dating may in the future prove an important link to the enigma. In 1955, Morris Swadesh introduced a mathematical method for determining the point in time at which two or more related languages had split – a method he called *glottochronology*.

This method has faced a great deal of criticism<sup>11</sup>, mainly because of a few basic assumptions such as the constant retention of basic vocabulary through time. The assumption is that any language retains approximately 81% of its basic vocabulary after 1000 years, and through comparing the differences in basic vocabulary between the related languages, the time at which they parted is calculated. Statistical methods for determining the age of languages has been rejected by most linguists.

Few linguists have spent their time contemplating solutions to the weaknesses of computational methods, leaving the science nearly stagnant during the 20<sup>th</sup> century. When computational phylogenetics were introduced in genetics, issues similar to those of the linguistic methods arose. Instead of refuting computational phylogenetic methods, geneticists started working their way around problems they faced. This led to revolutionizing methods in evolutionary biology. Interestingly, there are a great deal of similarities between evolutionary biology and genetic linguistics (Gray & Atkinson, 2005), and in recent years the question of applying computational methods for determining language relationship has enjoyed a renaissance. Unfortunately, there are still many who reject the methods, based on arguments against glottochronology<sup>12</sup>.

---

11 Beckwith (2004:200–210), Robbeets (2005:51), Gray & Atkinson (2003)

12 Cf. Gray & Atkinson (2003; 2005)

## Bibliography

Barnes, Gina L. (1999): *The Rise of Civilization in East Asia*. London: Thames and Hudson

Beckwith, Christopher I. (2004): *Koguryo: The Language of Japan's Continental Relatives: An Introduction to the Historical-Comparative Study of the Japanese-Koguryoic Languages*. Leiden: Brill

Beckwith, Christopher I. (2006): “The Ethnolinguistic History of the Early Korean Peninsula Region: Japanese-Koguryoic and other Languages in the koguryo, Paekche, and Silla kingdoms”. *Journal of Inner and East Asian Studies* 2: 33–65

Benedict, Paul K. (1990): *Japanese / Austro-Tai*. Ann Arbor: Karoma Publishers

Brown, Keith & Ogilvie, Sarah (2009): *Concise encyclopedia of languages of the world*. Amsterdam and Boston: Elsevier

Campbell, Lyle. (2004) *Historical Linguistics: An Introduction, 2<sup>nd</sup> edition*. Edinburgh: Edinburgh University Press

Diamond, Jared (1998): “Japanese roots”. *Discover Magazine*.  
<<http://discovermagazine.com/1998/jun/japaneseroots1455/>> (2010 -06-04)

Gray, R.D. & Atkinson, Q.D. (2003): “Language-tree divergence times support the Anatolian theory of Indo-European origin”. *Nature*, 426, 435–439

Gray, R.D. & Atkinson, Q.D. (2005): “Curious Parallels and Curious Connections—Phylogenetic Thinking in Biology and historical linguistics”, *Systematic Biology*, 54(4), 513–526

Kamei, T. (1969): *The Relationship of Japanese to the other Languages of East Asia*. Reprinted in T. Kamei, *Nihongo no keitōron no michi*. Tokyo: Yoshikawa-kōbundō, 1973

Karfunkel, Torbjörn. (2004): “Bayesian Inference using Markov Chain Monte Carlo in

Phylogenetic Studies". Essay written for the course in Markov Chains 2004

Labberton, D. van H. (1924): "Preliminary results of researches into the original relationship between the Nipponese and the Malay-Polynesian languages". *Journal of the Polynesian Society* 33:244–280

Martin, S. E. (1966): "Lexical evidence relating Korean to Japanese". *Language* 12.2, 185–251

Ottosson, I. & Ekholm, T. (2007): *Japans historia*. Falun: Scandbook

Robbeets, Martine Irma. (2004): "Belief or argument? The classification of the Japanese language". *Eurasia Newsletter* 8. Graduate School of Letters, Kyoto University

Robbeets, Martine Irma (2005): *Is Japanese related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden: Otto Harrassowitz

Robbeets, Martine Irma et al. (2006): Conference report of "The International Conference on the Language(s) of Koguryo and the Reconstruction of Old Korean and Neighboring Languages" held in Hamburg on September 23–24, 2005. *Journal of Inner and East Asian Studies* 2: 209–216

Sakiyama, Osamu (2001): "Kodai nihongo to oosutoronesiashogengo ni okeru ichikeitai no dougensei", *Bulletin of the Graduate School of Social and Cultural Studies*, Kyushu University. Vol. 7, 57–68

Shibatani, Masayoshi (1990): *The languages of Japan*. 8. ed., 2005, Cambridge: Cambridge University Press

von Siebold, Philipp Franz Balthazar. (1832): "Verhandeling over de afkomst der Japanners", *Verhandelingen des Bataviaasch Genootschaps* 13, 185–275

Solnit, David B. (1992): [review of] "Austro-Tai". *Language*, Vol. 68, No. 1

Svantesson, Jan-Olof. (2010-05-16): *Hmong-mienspråk*. <<http://www.ne.se/lang/hmong->>

mienspråk>

Swadesh, Morris. (1955): "Towards Greater Accuracy in Lexicostatistic Dating". *International Journal of American Linguistics* 21, 121–137

Tsutsui, William M. A. (2006): *Companion to Japanese History*. Blackwell Publishing, Blackwell Reference Online

Unger, J. Marshall. (2001): "Layers of Words and Volcanic Ash in Japan and Korea". *Journal of Japanese Studies*, Vol. 27, No. 1., 81–111