

Improving Speech to Text Services with Transcription Alignment

Ryan Crenny¹ Kaiser Pister²
Leelaludo

Abstract

Several transcription services report their own word error rates on public datasets but none report evaluations against real world video data such as public YouTube videos. This study aims to create an ideal transcript based on three different existing speech to text transcription services by using alignment and capitalizing on each services' strengths while also mitigating the others' weaknesses. Our generated transcript's accuracy was compared against human transcriptions using word error rate and word information loss. The results generally illustrate an increase in these metrics, which overall indicates the creation of a more accurate transcript.

1: Introduction:

Leelaludo's software allows for automated video translation via a speech to text transcription. That transcription is translated into a foreign language transcript before text to speech services recreate the original video with new audio. In order to prevent cascading errors in translation of the final project, the initial speech to text transcript must be as accurate as possible. Leelaludo has the option of using three different transcription services: Amazon Web Services Transcribe (AWS)³, YouTube Auto Generated Captioning (YouTube)⁴ and Azure Transcription⁵. Each of these services has its own strengths and weaknesses, so the goal of this project is to combine these services into an improved transcript.

2: Methods:

The first step is normalizing the transcripts into word lists from their JSON files, stripping punctuation, expanding contractions, and substituting colloquialisms.⁶ Then we "align" the transcripts on areas where they match. Alignment highlights the problematic areas of each transcripts where two or three of them differ. We then analyze and evaluate the three options to choose the best candidate with our "replacement" method. In order to simplify alignment, until this point all work is done on "stripped" transcripts i.e. all lowercase transcripts without any punctuation. The presence of punctuation is extremely important for future translation, so a "reconstruction" process is needed to add it back. Finally, the product's accuracy was tested on several videos where a near-perfect human transcript already existed by utilizing multiple quality

¹ rcrenny@gmail.com

² <https://www.audiodub.app>

³ <https://aws.amazon.com/transcribe/>

⁴ <https://developers.google.com/youtube/v3/docs/captions>

⁵ <https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/>

⁶ <https://pypi.org/project/jiwer/>

metrics. Our aligned transcript vs. the human transcript's score performed comparably or better than each individual transcription services' score.

3: Alignment:

The three transcripts from each service are input as lists of words. "Root" words are found within these lists, on the condition that the word appears once per transcript, so we know it is a unique word. We ensure alignment in the roots by validating that the roots appear in the same order in each list, removing those which are out of order. Then the list of roots is iterated through and "working segments" are created and the output list appends a recursive call of the align function with each working set minus the final word. This recursive call allows the alignment problem to be split up into many smaller problems, until it gets to the point where two working sets have zero words in common, so a tuple is added to the output list on these non overlapping sections.

4: Replacement:

The output from alignment is passed into replacement. This input is a list of words and tuples. Each tuple represents a difference in the transcription services' output. We iterate through the list, analyzing each tuple to choose the best possible option. We evaluate each tuple with three heuristics. First If two of the three tuple values are equal to each other, then that shared value is chosen. Second, in the case where all the tuple values are only made up of one word, but that word is different for each transcription, we compare the words with the Jaro⁷ similarity metric for any given two words. Jaro similarity is calculated based on the number of matching characters, the string lengths, and the number of transpositions, which are matching characters that are in different orders. Using this metric helps us decide which two words are most alike, allowing us to rule out at least one of the options and default to one of the two alike ones. Third, if none of the tuple values equal each other and at least one value is longer than one word, the tuple is passed into our reconstruction method.

```
1 for value in list:
2     if value is tuple:
3         if 2 tuple values equal each other:
4             append(matching pair value)
5         elif all tuple values are one word:
6             append(handleSingleWord(value))
7         else:
8             append(reconstruct(value))
```

⁷ Cohen, W. W.; Ravikumar, P.; Fienberg, S. E. (2003). <https://www.cs.cmu.edu/afs/cs/Web/People/wcohen/postscript/kdd-2003-match-ws.pdf>. *KDD Workshop on Data Cleaning and Object Consolidation*. 3: 73–8.

```
# Pseudocode of Replacement function
Examples:
Matching Pair: (['trickling'], ['trickling'], ['trickle', 'in'])
Single Word: (['but'], ['that'], ['there']) #appends "that"
handleOther: (['roll', 'logo'], ['look', 'at', 'them', 'yeah', 'i'],
['look', 'at', 'them', 'i']) #appends "look at them i"
```

This method attempts to align two of the three values in order to find the best match possible. Following the same logic as the Matching Pair heuristic, two similar values more likely predict the correct value. Assuming we find an alignment between two of the three values, we further replace the lists to arrive at a single string. If there is no match against the values, we finally default to one of the three services. The idea behind this approach is that these tuple values exist because alignment tells us they have no matches between the three of them, but there could still be enough similarity between any given two that they could be further aligned, as to give us a more accurate output.

5: Reconstruction:

The output from replacement is meant to have the best lexical options, but punctuation must still be reconstructed in order to have the best possible final transcript. The initial input parameters to this method are the replaced transcript, which is all lowercase with no punctuation, and one of the original punctuated transcripts. Due to this difference in punctuation, there is handling for stripping punctuation and capitalization everywhere it is necessary. The beginning is almost exactly the same as alignment with a list of root words created between the two lists. These root words help us create a list of segments of the same length. We iterate through these lists of segments, and if the lengths of the two segments are equal, then the case is simple and `handleSameLength` is called. This method iterates through the segments, appending equal words, and adding punctuation where it is needed. If the segments' lengths are not equal then the same root finding process as above is called strictly on each segment. If the length of these new roots is greater than one, then the method is called recursively on strictly these two segments. Otherwise, the shorter of the two segments has null values added to it, in order to equal the length of the longer segment, allowing `handleSameLength` to be called on these two segments.

6: Evaluation:

In order to provide the most robust and indicative results, Word Error Rate (WER) and Word Information Loss (WIL)⁸ metrics are both presented below. WER is commonly used in automatic speech recognition assessment. Morris et al. argue for a set of more robust metrics for ASR, such as WIL which measures the proportion of word information communicated by utilizing an equation that links information with Pearson's Large Sample Statistic and retains only dominant terms. Their paper reports that the two metrics give similar scores for low error

⁸ Morris, Andrew Cameron / Maier, Viktoria / Green, Phil (2004): "From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition", In *INTERSPEECH-2004*, 2765-2768.

rates, but on many real world applications, the scores given by each metric begin to differ significantly.

Using the above criteria, we ask the following three questions:

1. Which transcription service is best to use as the default option?
2. Which transcription service provides the best punctuation?
3. How do our generated transcript's scores compare to each individual transcription service?

7: Results:

1. Which transcription service is best to use as the default option?

Key:

Matching Pair: two of the three tuple values equal each other, so that is chosen

One Word: all three tuple values are one word, so handleOneWord is called, which chooses the best word via a likeness score

Further Align: reconstruction is called on the tuple values

Default: The values are too disparate, so a default value is chosen

v1: <https://www.youtube.com/watch?v=bia84-cftNc>

v2: <https://www.youtube.com/watch?v=W8JBPYJNkV8>

v3: <https://www.youtube.com/watch?v=TK-zAsfwF8c>

v4: <https://www.youtube.com/watch?v=IvmHU5Hcs2k>

	v1	v2	v3	v4
Matching Pair	34 (81%)	85 (57%)	99 (84%)	25 (89%)
One Word	0	2 (1%)	1 (1%)	1 (3%)
Further Align	6 (14%)	39 (26%)	9 (7%)	0
Default	2 (4%)	23 (15%)	8 (7%)	2 (7%)

This table above shows the distribution of each heuristic we developed.

This data below is all without punctuation, since this experiment is strictly analyzing the lexical options.

AZURE Default					
	v1	v2	v3	v4	Avg
First Condit	14.91	34.66	10.83	5.91	16.5775
First Condit+Single Word	14.91	34.54	10.83	6.21	16.6225
First Condt+Further Align	12.72	31.64	11.08	5.91	15.3375

Final	12.72	31.52	11.08	6.21	15.3825
PyTube Default					
	v1	v2	v3	v4	Avg
First Condit	13.43	29.79	11.27	6.06	15.1375
First Condit+Single Word	13.43	29.79	11.38	6.06	15.165
First Condt+Further Align	13.43	30.09	10.91	6.06	15.1225
Final	13.43	30.09	11.02	6.06	15.15
AWS Default					
	v1	v2	v3	v4	Avg
First Condit	12.05	31.57	12.52	5.75	15.4725
First Condit+Single Word	12.05	31.45	12.52	5.75	15.4425
First Condt+Further Align	13.78	30.04	11.47	6.06	15.3375
Final	13.78	30.04	11.58	6.06	15.365

*all WIL scores

Clearly none of the services provide a definitive edge, but we decided to choose YouTube because of its better overall average score.

2. Which transcription service provides the best punctuation?

As stated above, punctuation is extremely important when it comes to translation. A general rule is that more punctuation leads to a better translation. Lack of punctuation can entirely change the meaning of a phrase, such as:

The man dropped the bullet in his mouth. (vs) The man dropped, the bullet in his mouth.
Let's eat Grandma. (vs) Let's eat, Grandma

Punctuation can obviously affect sentence meaning, and therefore lead to a much worse translation. YouTube does not provide punctuation, and AWS has trouble handling background noise and music, which often leads to transcriptions of nonsensical phrases, or in the case of music, just the word “yeah” being transcribed over and over again. With all this in mind, we decided to use Azure as our punctuation base.

3. How do our generated transcript's scores compare to each individual transcription service?

Note: our human transcripts were based on AWS, resulting in a bias towards AWS's scores, specifically on the punctuated transcripts

Without Punctuation

WIL	v1	v2	v3	v4
Aligned	13.73	29.54	10.88	6.06
AWS	13.43	29.06	14.88	8.02
Azure	18.99	36.19	13.3	8.02
YouTube	15.03	29.62	13.9	8.02
WER	v1	v2	v3	v4
Aligned	11.32	24.78	9.4	3.77
AWS	10.68	25.33	14.53	5.03
Azure	13.46	27.87	10.43	5.03
YouTube	12.82	22.95	11.58	5.03

With Punctuation

WIL	v1	v2	v3	v4
Aligned	24.99	33.22	16.70	15.59
AWS	17.27	29.15	18.31	15.3
Azure	29	40.18	18.27	17.24
YouTube	36.07	43.1	27.55	22.78
WER	v1	v2	v3	v4
Aligned	17.73	25.59	12.84	8.8
AWS	12.82	24.82	16.52	8.8
Azure	19.65	29.97	13.2	10.06
YouTube	25.21	31.51	19.48	13.05

8: Conclusion:

Generally, the strategy employed here was to divide the problem into many smaller parts. This approach was used in alignment as well as the reconstruction. Initially the replacement method only handled the Matching Pair, and otherwise used a default option, however after completing the reconstruction method, further handling was implemented for single word options and further alignment. Adding both of these conditions provides more robust handling and avoids having to use a default option. In handling this problem several areas were specialized to

account for strengths and weaknesses of the transcripts. The reconstruction method requires one of the transcription services to be used as a “base” punctuated transcript, and we elected to use Azure for this because it tends to use more punctuation than the others. There are also going to be instances within replacement where the three options are so different from each other that none of our handling applies, in these cases a default option must be chosen and upon analysis we decided that YouTube was the best option. With all of this implemented, we found it best to separate our score comparisons into two categories, with and without punctuation. Ostensibly, analysis with punctuation seems to perform poorly, but one must keep in mind that the human transcripts were created with AWS as a base, which explains why AWS outperforms us for both metrics. On the other hand, without punctuation, the aligned transcript performs better, specifically with the WIL metric, where our transcript outperforms the services on three of the four videos. Some of these improvements may only be marginal, but it is still important to highlight that our approach to the problem did yield the desired results, small as that improvement may be. Potential improvements to our method could be a more robust further alignment condition call within the replacement method. Currently this call utilizes the reconstruction method, which was created with specific default weights to choose values, so it could be improved by creating a method that is specifically designed for this condition.

References:

- AWS Transcription: <https://aws.amazon.com/transcribe/>
- Youtube Auto Generated Captioning:
<https://developers.google.com/youtube/v3/docs/captions>
- Azure Transcription:
<https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/>
- JiWER: Similarity measures for automatic speech recognition evaluation:
<https://pypi.org/project/jiwer/>
- Cohen, Ravikumar, Fienberg(2003) “A Comparison of String Metrics for Matching Names and Records”, In IIWEB'03: Proceedings of the 2003 International Conference on Information Integration on the WebAugust 2003 Pages 73–78
- Morris, Andrew Cameron / Maier, Viktoria / Green, Phil (2004): "From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition", In *INTERSPEECH-2004*, 2765-2768.

Appendix:

v1:

value (Azure, AWS, YouTube)	Condition	Aligned Value	Human
(['blog', 'vlog', 'vlog'])	Matching Pair		vlog
([], [], ['uh'])	Matching Pair		[]
(['2', '1/2'], ['2.5'], ['two', 'and', 'a', 'half'])	DEFAULT		2.5
(['a', 'blog'], ['the', 'blog'], ['the', 'vlog'])	FURTHER ALIGN	['the', 'vlog']	The vlog
(['would', 'see'], ['would', 'at', 'sea'], ["wouldn't", 'see'])	FURTHER ALIGN	["wouldn't", 'see']	Out at sea
(['out', 'at', 'sea'], ['owed', 'etc'], ['out', 'at', 'sea'])	Matching Pair		Out at sea
([], ['oh'], ['oh'])	Matching Pair		[]
([], [], ['uh'])	Matching Pair		[]
(['going', 'to'], ['gonna'], ['gonna'])	Matching Pair		Going to
(['patreon.com/route', 'so'], ['patreon', 'dot', 'com', 'slash', 'roots', 'of'], ['patreon.com', 'roots', 'of'])	FURTHER ALIGN	patreon.com, 'roots', 'of']	
(['as'], ['because'], ['as'])	Matching Pair		
(['i'], ['um', 'uh'], ['um', 'i', 'uh'])	DEFAULT		
(['i', 'i'], ['i'], ['i', 'i'])	Matching Pair		
(["it's"], ['has'], ["it's"])	Matching Pair		
([], ['uh'], [])	Matching Pair		
(['while'], ['well'], ['well'])	Matching Pair		
(['of'], ['of'], [])	Matching Pair		
(['trickling'], ['trickling'], ['trickle', 'in'])	Matching Pair		

(['crew'], ['group'], ['crew'])	Matching Pair		
(['will'], [], [])	Matching Pair		
(['and'], [], ['and'])	Matching Pair		
(['going', 'to'], ['gonna'], ['gonna'])	Matching Pair		
(['going', 'to'], ['gonna'], ['gonna'])	Matching Pair		
([], ['we'll'], ['we'll'])	Matching Pair		
([], [], ['uh'])	Matching Pair		
(['thanks', 'thanks'], ['uh', 'thanks', 'thanks'], ['uh', 'thanks', 'thanks'])	Matching Pair		
([], ['uh'], ['uh'])	Matching Pair		
([], ['don't'], ['don't'])	Matching Pair		
(['has'], ['that's'], ['that's'])	Matching Pair		
(['cause'], ['because'], ['because'])	Matching Pair		
([], ['uh'], [])	Matching Pair		
(['an'], [], [])	Matching Pair		
([], [], ['uh'])	Matching Pair		
(['try', 'try', 'to', 'keep', 'things', 'keep', 'things'], ['try', 'try', 'to', 'keep', 'things', 'keep', 'things'], ['i', 'try', 'to', 'try', 'to', 'keep', 'things', 'keep', 'things'])	Matching Pair		
(['alright'], ['yeah', 'all', 'right'], ['all', 'right'])	FURTHER ALIGN	all, 'right']	
(['gotta', 'get'], ['got', 'to', 'go'], ['gotta', 'go'])	FURTHER ALIGN	gotta', 'go']	
(['sailing'], ['sail', 'on'], ['sailing'])	Matching Pair		
([], [], ['uh'])	Matching Pair		

([], ["it'll", 'take', 'me'], ["it'll", 'take', 'me'])	Matching Pair		
(['gotta'], ['i', 'got', 'to'], ['got', 'to'])	FURTHER ALIGN	got', 'to']	

v2:

value (Azure, AWS, YouTube)	Condition	Aligned Value
([], ['mm', 'mm', 'hmm', 'mm', 'yeah', 'yeah', 'yeah', 'you', 'want', 'to', 'yeah', 'yeah', 'was'], ['so', 'so', 'this'])	DEFAULT	
(['haul'], ['hall'], ['hall'])	Matching Pair	
(['type', 'one', 'unsuccessful', 'and'], ['okay', '10', 'successful', 'we'll', 'be', 'too', 'close', 'to', 'this', 'one', 'and'], ['10th', 'one', 'and', 'successful', 'a', 'little', 'bit', 'too', 'close', 'to', 'this', 'one', 'and'])	FURTHER ALIGN	['10th', 'one', 'and', 'successful', 'a', 'little', 'bit', 'too', 'close', 'to', 'this', 'one', 'and']
(['mean', "it's"], ['moved', 'to'], ['made', 'it', 'to'])	FURTHER ALIGN	['made', 'it', 'to']
(['going', 'to'], ['going', 'to'], ['gonna'])	Matching Pair	
(['have'], ['have'], ["i've"])	Matching Pair	
(['dhamma', 'malawi', 'overpass', 'and', 'a'], ['drama', 'melayu', 'vipassana'], ['dhamma', 'malayu', 'vapasana'])	DEFAULT	
(['this', 'is'], ["it's"], ["it's"])	Matching Pair	
(['ok'], ['okay'], ['okay'])	Matching Pair	

(['at', 'of', 'apocynum'], ['out', 'of', 'a', 'passive'], ['at', 'a', 'vipassana'])	FURTHER ALIGN	at', 'a', 'vipassana']
(['sane'], ['scene'], ['sane'])	Matching Pair	
(['dhamma', 'malaya', 'but', 'pasnau', 'center', 'is'], ['dam', 'um', 'alaya', 'vipassana', 'centre', 'just'], ['dhamma', 'malaya', 'vapasana', 'center', 'this', 'is'])	FURTHER ALIGN	['dhamma', 'malaya', ' 'vapasana', 'center', 'this', 'is']
(['3'], ['three'], ['three'])	Matching Pair	
(["it's"], ['it', 'was'], ['it', 'was'])	Matching Pair	
([], ['i'll'], [])	Matching Pair	
(['surpassing', 'centers'], ['vipassana', 'centres'], ['vipassana', 'centers'])	FURTHER ALIGN	['vipassana', 'centers']
([], ['here'], ['here'])	Matching Pair	
(['or'], ['your'], ['your'])	Matching Pair	
(['your'], ['your'], ['you'])	Matching Pair	
([], ['yeah', 'and', 'then'], ['and', 'then'])	FURTHER ALIGN	['', 'and', 'then']
(['died'], ['die'], ['dine'])	SINGLE WORD	
(['brings'], ['bring'], ['brings'])	Matching Pair	
(['drives'], ['drives'], ['dries'])	Matching Pair	
(["there's"], ['is'], ["there's"])	Matching Pair	
(['you', 'started', 'having'], ["you'd", 'sort', 'of', 'have'], ['you', 'sort', 'of', 'have'])	FURTHER ALIGN	['you', 'sort', 'of', 'have']

(['or'], ['we'], ['well', 'you'])	DEFAULT	
([], ['yeah'], [])	Matching Pair	
(['we', 'died'], ['we're', 'dying'], ['we', 'die'])	FURTHER ALIGN	['we', 'die']
(['in'], ['on'], ['on'])	Matching Pair	
([], ['you', 'have'], ['now', 'we', 'have'])	FURTHER ALIGN	['now', 'we', 'have']
(['me', 'low'], ['melo'], ['milo'])	DEFAULT	
(['that', 'review'], ['every', 'meal'], ['every', 'meal'])	Matching Pair	
(['you'], [], ['you'])	Matching Pair	
(['the'], ['the'], ['that'])	Matching Pair	
(['and', 'the', 'company'], ['when', 'you', 'come', 'and'], ['and', 'then', 'you', 'come', 'and'])	DEFAULT	
(['under', 'mat'], ['under', 'now'], ['on', 'your', 'mat'])	FURTHER ALIGN	[[['under', 'now'], ['on', 'your', 'mat']]]
(['here'], ['you', 'hear', 'the'], ['and', 'hear', 'the'])	FURTHER ALIGN	['and', 'hear', 'the']
(['reach'], ['huh', 'yeah', 'huge'], ['huh', "you're", 'each'])	FURTHER ALIGN	['huh', "you're", 'each']
([], ['those'], ['nice', 'little'])	DEFAULT	
([], ['the', 'floor', 'here'], ['the', 'floor', 'here'])	Matching Pair	
(['you'], ['do', 'you'], ['let', 'me'])	FURTHER ALIGN	['do', 'you']
([], [], ['in', 'here'])	Matching Pair	
(['ceiling'], ['yeah', 'what', 'you'], ['ceiling', 'man', 'what', 'you'])	DEFAULT	

(['wanna'], ['want', 'to'], ['want', 'to'])	Matching Pair	
([], [], ['is'])	Matching Pair	
([], ['you'], ['because', 'you'])	FURTHER ALIGN	['because', 'you']
(['clock'], ['cloth', "it's"], ['clock', "it's"])	DEFAULT	
(['this', 'they'], ['you', 'so', 'you'], ['you', 'so', 'you'])	Matching Pair	
(['if'], ['did'], ['if'])	Matching Pair	
(['of'], ['with'], ['with'])	Matching Pair	
(['more', 'more', 'painful', 'salaria'], ['more', 'and', 'more', 'pain', 'cell', 'area'], ['more', 'more', 'painful', 'the', 'cell', 'area'])	FURTHER ALIGN	['more', 'more', 'painful', 'the', 'cell', 'area']
(["it's", 'pretty'], [], ["it's", 'pretty'])	Matching Pair	
(['my', 'cell', 'cell', 'number'], ['myself', 'cell', ""], ['my', 'cell', 'cell', 'number'])	Matching Pair	
([], [], ["it's"])	Matching Pair	
(['voluntary', 'solitaire', 'e', 'confinement', 'voluntary', 'solitary', 'confinement', 'coming'], ['voluntary', 'solitary', 'confinement', 'voluntary', 'solitary', 'come', 'fine', 'man', 'come', 'in'], ['voluntary', 'solitary', 'confinement', 'voluntary', 'solitary', 'confinement', 'come', 'in'])	FURTHER ALIGN	voluntary', 'solitary', 'confinement', 'voluntary', 'solitary', 'confinement', 'come', 'in']
(['set'], ['sit'], ['you', 'sit'])	FURTHER ALIGN	['you', 'sit']
(['go'], [], ['go'])	Matching Pair	

(['going'], ['going'], ['go', 'on'])	Matching Pair	
(['paying'], ['painting'], ['pain'])	SINGLE WORD	
(["don't", 'get', 'anywhere'], ['know', 'okay', 'so'], ["don't", 'get', 'anywhere', 'so'])	DEFAULT	
(['you'], [], ['you'])	Matching Pair	
(['dustpan'], ['dust', 'pan'], ['dustpan'])	Matching Pair	
(['foam'], ['boom', 'mm', 'hmm'], ['boom'])	DEFAULT	
([], ['looks', 'like'], ['looks', 'like'])	Matching Pair	
(['so'], ['except'], ['except'])	Matching Pair	
(['christian', 'location', 'map'], ['pushing', 'the', 'meditation', 'mat', 'and'], ['cushion', 'a', 'meditation', 'mat', 'a'])	FURTHER ALIGN	['cushion', 'a', 'meditation', 'mat', 'a']
(['place'], ['pleased'], ['place'])	Matching Pair	
(['even', 'a'], ['you', 'in', 'the'], ["there's", 'even', 'a'])	FURTHER ALIGN	["there's", 'even', 'a']
([], ['and', 'the'], ['and', 'the'])	Matching Pair	
(['between', 'but', 'this'], ['you', 'gotta', 'have', 'a', 'clean', 'but', 'it's', 'a', 'decent'], ['have', 'a', 'clean', 'butt', 'it's', 'a', 'decent'])	FURTHER ALIGN	have', 'a', 'clean', 'butt', "it's", 'a', 'decent']
(['to'], ['to'], ['too'])	Matching Pair	
(['my', 'privacy', 'and'], ['mom', 'privacy', 'but'], ['non-privacy', 'oh', 'yeah'])	FURTHER ALIGN	['mom', 'privacy', 'but']
(['which', 'i', "didn't"], ['okay', 'we', 'can'], ['which', 'i', "didn't"])	Matching Pair	

(['that', 'they', 'may', 'be', 'hide'], ['their', 'their', 'meaning', 'behind'], ['there', 'they', 'mean', 'behind'])	FURTHER ALIGN	['there', 'they', 'mean', 'behind']
(['just', 'needs', 'to', 'be', 'more', 'open'], ['sure'], [])	DEFAULT	
(['or', 'or'], ['or', 'or'], ['or', 'or'])	Matching Pair	
(['there', 'is', 'a'], ['there', 'is', 'a'], ["there's", 'the'])	Matching Pair	
(['they', 'are'], ["they're"], ["they're"])	Matching Pair	
(['the'], ['the'], ['a'])	Matching Pair	
(['4:00'], ['four'], ['four'])	Matching Pair	
(['4:30'], ['4', '30'], ['4', '30'])	Matching Pair	
(['meditating'], ['meditate', 'in'], ['meditate', 'in'])	Matching Pair	
(['4:30'], ['4', '30'], ['4', '30'])	Matching Pair	
(['6:30'], ['6', '30'], ['6', '30'])	Matching Pair	
(['and'], ['when'], ['and'])	Matching Pair	
(['8'], ['eight'], ['8'])	Matching Pair	
(['1:00'], ['one'], ['one'])	Matching Pair	
(['5:00'], ['five'], ['five'])	Matching Pair	
(['tea'], ['t'], ['tea'])	Matching Pair	
(['6'], ['six'], ['six'])	Matching Pair	

(['goenka'], ['glinka'], ['goenka'])	Matching Pair	
(['the', 'pasta'], ['for', 'passing'], ['with', 'pastor'])	DEFAULT	
(['and', 'dhamma', 'and'], ['and', 'drama', 'and'], ['and', 'dhamma', 'and'])	Matching Pair	
(['9:00'], ['nine'], ['nine'])	Matching Pair	
(['and'], ['then'], ['and'])	Matching Pair	
(['4:00'], ['four'], ['four'])	Matching Pair	
([], ['you'], [])	Matching Pair	
(['day', 'after', 'day', 'after', 'day'], ['day', 'after', 'day', 'after', 'day'], ['day', 'after', 'day', 'after', 'day'])	Matching Pair	
(['anapana'], ['and', 'upon', 'a'], ['anapana'])	Matching Pair	
(['so'], [], [])	Matching Pair	
(['going', 'to'], ['gonna'], ['going', 'to'])	Matching Pair	
([], ['yeah'], [])	Matching Pair	
(['pacyna', 'center'], ['vipassana', 'centre'], ['vapasana', 'center'])	FURTHER ALIGN	['vapasana', 'center']
(['there'], ['there'], ['their'])	Matching Pair	
(['there'], ["they're"], ["they're"])	Matching Pair	
(['roll', 'logo'], ['look', 'at', 'them', 'yeah', 'i'], ['look', 'at', 'them', 'i'])	FURTHER ALIGN	['look', 'at', 'them', ' ', 'i']
(['yeah', 'yeah'], ['yeah'], ['yeah'])	Matching Pair	
([], ['at'], ['at'])	Matching Pair	

(['these'], ['these'], ['he's'])	Matching Pair	
(['i'll', 'phone'], ['i', 'have', 'fallen'], ['i', 'have', 'fallen'])	Matching Pair	
(['liberty'], ['love', 'with', 'these'], ['love', 'with', 'his'])	FURTHER ALIGN	['love', 'with', 'his']
(['is', '6'], ['these', 'six'], ['he's', 'six'])	FURTHER ALIGN	['he's', 'six']
(['we', 'are', 'we'], ['we', 'we'], ['we', 'uh', 'we'])	DEFAULT	
(['else', 'is'], ['else', 'is'], ['else's'])	Matching Pair	
(['week', 'week', 'give'], ['we', 'can', 'you', "can't", 'leave'], ['we', "can't", 'leave'])	FURTHER ALIGN	['we', ", ", "can't", 'leave']
(['their', 'moms'], ['mom's'], ['the', "mom's"])	FURTHER ALIGN	['the', "mom's"]
(['so'], ['so'], ['sir'])	Matching Pair	
(['in', 'my', 'life', 'terpin'], ['him', 'alive', 'yeah'], ['them', 'alive'])	DEFAULT	
(['wanna', 'say', 'who', 'was', 'the', 'cia', 'to'], ['want', 'to', 'say', 'who', 'wants', 'to', 'see', 'a', 'as', 'a'], ['want', 'to', 'say', 'who', 'who', 'wants', 'to', 'say', 'hi', 'to', 'the'])	FURTHER ALIGN	['want', 'to', 'say', 'who', 'who', 'wants', 'to', 'say', 'hi', 'to', 'the']
(['for', 'the'], ['your', 'little'], ["you're", 'the'])	FURTHER ALIGN	["you're", 'the']
(['oh', 'yeah', 'oh', 'yeah', 'hi', 'some', 'yeah'], ['oh', 'yeah', 'oh', 'yeah', 'nice', 'um', 'yeah'], ['oh', 'yeah', 'oh', 'yeah', 'handsome', 'yeah'])	DEFAULT	
(['so'], ["it's", 'a'], ["it's", 'a'])	Matching Pair	

(['to', 'adopter', 'yeah', 'who', 'wants', 'who', 'wants', 'to'], ['to', 'a', 'doctor', 'yeah', 'who', 'wants', 'who', 'wants', 'to'], ['to'])	FURTHER ALIGN	['to', 'a', 'doctor', 'yeah', 'who', 'wants', 'who', 'wants', 'to']
([], [], ['her', 'yeah', 'who', 'wants', 'to', 'watch', 'the', 'dog'])	Matching Pair	
(['cause', "we've"], ['because', 'people'], ['because', "we've"])	FURTHER ALIGN	['because', "we've"]
(['2'], ['two'], ['two'])	Matching Pair	
([], [], ['and'])	Matching Pair	
(['to', 'keep', 'him', 'alive', 'like', "that's", 'not', 'a', "", 'oh'], ['to', 'kill', 'my', 'life', 'oh', 'glad', 'to', 'give', 'you', 'oh'], ['to', 'keep', 'them', 'alive', 'oh', 'oh'])	DEFAULT	
(['oh', 'yeah'], ['yeah', 'yeah', 'oh', 'yeah', 'oh', 'bless', 'you', 'oh', 'it', 'was'], ['yeah', 'oh', 'yeah', 'oh', 'pleasure', 'no'])	DEFAULT	
(['but'], [], ['well'])	DEFAULT	
(['better', 'two', 'months', 'in', 'in', 'none'], ['bad', 'two', 'months', 'and', 'then', 'and', 'then', 'okay', 'then', "we'll"], ['better', 'too', 'much', 'than', 'that', 'okay', 'and', 'then', 'you'])	FURTHER ALIGN	better', 'too', 'much', 'than', 'that', 'okay', 'and', 'then', 'you']
(['no', 'no', 'movie'], ['you', 'want', 'a', 'little', 'bit', 'no', 'okay', "that's", 'it'], ['okay', 'you', 'want', 'a', 'little', 'bit', 'nope'])	DEFAULT	
(['want', 'to'], ['wanna'], ['want', 'to'])	Matching Pair	

(['at', 'home', 'there', 'we', 'go', 'i'], ['and', 'him', 'oh', 'sure', 'there', 'we', 'go', 'oh', "i'll"], ['with', 'them', "i'll"])	FURTHER ALIGN	with', 'them', "i'll"]
(['georgia', 'said', 'yesterday', 'yesterday', 'versus'], ['we', 'were', 'just', 'saying', 'yesterday', 'did', 'a', 'video', 'yesterday'], ['uh', 'we', 'were', 'just', 'saying', 'yesterday', 'we', 'did', 'a', 'video', 'yesterday'])	FURTHER ALIGN	['uh', 'we', 'were', 'just', 'saying', 'yesterday', 'we', 'did', 'a', 'video', 'yesterday']
([], ['them'], ['them'])	Matching Pair	
(['in', 'coca', 'cola'], ['from', 'coca', 'cola'], ['in', 'coca-cola'])	FURTHER ALIGN	['from', 'coca', 'cola']
(['him'], ['them'], ['them'])	Matching Pair	
(["mcdonald's"], ['mcdonalds'], ["mcdonald's"])	Matching Pair	
(['yeah', 'please', 'yeah', 'yeah', 'yeah', 'yeah', 'agreement', 'building', 'the'], ['fish', 'and', 'chips', 'yeah', 'yeah', 'yeah', 'and', 'the', 'agreement', 'building', 'the'], ['the'])	FURTHER ALIGN	['fish', 'and', 'chips', 'yeah', 'yeah', 'yeah', 'and', 'the', 'agreement', 'building', 'the']
(['keep', 'this', 'is', 'this', 'is', 'this', 'is'], ['you', 'can', 'make', 'an', 'idea', 'this', 'is', 'this', 'is', 'this', 'is'], ['this', 'is', 'this', 'is', 'this', 'is'])	DEFAULT	
(['adama'], ['a', 'dime', 'a'], ['the', 'dhamma'])	DEFAULT	
(['sad', 'like', 'during', 'the'], ['the', 'time', 'my', 'father', 'yeah', 'oh', 'yeah', 'it', 'looks', 'like', 'during', 'the', 'course', 'went', 'out', 'of', 'them', 'mm', 'workers', 'oh'], ['my', 'father'])	DEFAULT	

(['with', 'you', 'marcus', 'he'], ['miss', 'democracy'], ['with', 'democracy'])	FURTHER ALIGN	['with', 'democracy']
(['is'], [], ['this', 'is'])	FURTHER ALIGN	['this', 'is']
(['big', 'guy'], ['big', 'guy', 'big', 'guy'], ['big', 'guy'])	Matching Pair	
(['him'], ['them'], ['him'])	Matching Pair	
([], ['all', 'right'], [])	Matching Pair	
([], ['yeah', 'yeah', 'yeah', 'yeah', 'right', 'yeah', 'yeah'], ['so', 'you'])	DEFAULT	

v3:

value (Azure, AWS, YouTube)	Condition	Aligned Value
(['surfshark'], ['surf', 'shark'], ['surf', 'shark'])	Matching Pair	
(['multi', 'leg'], ['multi', 'legged'], ['multi-leg'])	FURTHER ALIGN	['multi', 'legged']
(['and'], ['and'], ['i'm'])	Matching Pair	
(['a380'], ['a', '3', '80'], ['a380'])	Matching Pair	
(['around'], ['a', 'round'], ['around'])	Matching Pair	
(['covid'], ['covid'], ['covert'])	Matching Pair	
(['to'], ['your'], ['to', 'your'])	DEFAULT	
(['red'], ['read'], ['red'])	Matching Pair	
([], ['mm'], [])	Matching Pair	
(['9:45'], ['9:45'], ['9', '45'])	Matching Pair	
(['3'], ['three'], ['3'])	Matching Pair	
([], [], ['it'])	Matching Pair	
(['kris'], ['chris'], ['chris'])	Matching Pair	
(['but'], ['that'], ['there'])	SINGLE WORD	
(['9:55', 'am'], ['9:55', 'a.m'], ['9', '55', 'am'])	FURTHER ALIGN	[['9:55', 'a.m']
(['9:45'], ['9', '45'], ['9', '45'])	Matching Pair	
(['planeload'], ['plane', 'load'], ['plane', 'load'])	Matching Pair	

(['katie', 'four'], ['cater', 'for'], ['cater', 'for'])	Matching Pair	
([], ['okay'], [])	Matching Pair	
(['jamun'], ['jamon'], ['jamun'])	Matching Pair	
(['donut'], ['doughnut'], ['donut'])	Matching Pair	
([], ['it's'], [])	Matching Pair	
(['and'], [], ['and'])	Matching Pair	
(['is'], ['is'], ['is', 'is'])	Matching Pair	
(['this', 'this'], ['this', 'uh', 'this'], ['this', 'uh', 'this'])	Matching Pair	
(['and', 'log'], ['and', 'log'], ['analog'])	Matching Pair	
(['really', 'the', 'company'], ['really', 'that', 'company'], ['really', 'though', 'i', "can't", 'really'])	FURTHER ALIGN	['really', 'that', 'company']
(['with'], ['of'], ['with'])	Matching Pair	
(['tickets'], ['tickets'], ['e-tickets'])	Matching Pair	
([], ['professional', 'okay', 'yeah', "we'll", 'go', 'there'], [])	Matching Pair	
(['surfshark', 'and', 'award-winning'], ['surf', 'shark', 'an', 'award', 'winning'], ['surf', 'shark', 'an', 'award-winning'])	FURTHER ALIGN	['surf', 'shark', 'an', 'award-winning']
(['surfshark'], ['surf', 'shark'], ['surfshark'])	Matching Pair	
(['wi-fi'], ['wifi'], ['wi-fi'])	Matching Pair	
(['whitelisting'], ['white', 'listing'], ['white', 'listing'])	Matching Pair	
(['industry', 'leading'], ['industry', 'leading'], ['industry-leading'])	Matching Pair	
(['it'], ["it's"], ['it'])	Matching Pair	
(['surfshark'], ['surf', 'shark'], ['surf', 'shark'])	Matching Pair	
(['vala'], ['voila'], ['voila'])	Matching Pair	
(['surfshark'], ['surf', 'shark'], ['surfshark'])	Matching Pair	
(['surfshark', 'dot', 'deals', 'forward'], ['surf', 'shark', 'dot', 'deals', 'four'], ['surfshark.deals', 'forward'])	FURTHER ALIGN	[(['surf', 'shark', 'dot', 'deals', 'four'])
(['wing'], ['winging'], ['winging'])	Matching Pair	
([], [], ['percent'])	Matching Pair	
(['surfshark'], ['surf', 'shark'], ['surfshark'])	Matching Pair	
([], ['mm'], [])	Matching Pair	
(['off', 'to'], ['off', 'to'], ['after'])	Matching Pair	
(['a380'], ['a', '380'], ['a380'])	Matching Pair	
(['business', 'cards'], ['okay', 'and', 'um', 'great', 'to', 'them', 'that', 'business'], [])	DEFAULT	

(['a', '380', 'hours'], ['a', '380', 'ours'], ['a380', 'ours'])	FURTHER ALIGN	['', 'a380', 'ours']
(['day'], ['deck'], ['deck'])	Matching Pair	
(['a380'], ['a', '380'], ['a380'])	Matching Pair	
(['there'], ['they're'], ['they're'])	Matching Pair	
(['seeds'], ['seats'], ['seats'])	Matching Pair	
([], [], ['i', 'have'])	Matching Pair	
(['hello', 'vivian', 'thank', 'you', 'hello', 'hello', 'welcome', 'back', 'thank', 'you'], ['yeah', 'yeah', 'like', 'that', 'thank', 'you', 'yeah', 'right', 'right', 'hello', 'thank', 'you', 'thank', 'you'], ['hello', 'thank', 'you'])	DEFAULT	
(['do'], ['yeah', 'yeah', 'yeah', 'is', 'it'], ['do'])	Matching Pair	
(['a'], ['a'], ['us', 'in', 'the'])	Matching Pair	
(['ok'], ['okay'], ['okay'])	Matching Pair	
(['pictures'], ['a', 'picture'], ['a', 'picture'])	Matching Pair	
([], ['the'], ['the'])	Matching Pair	
(['in', 'delhi'], ['so', 'how', 'is', 'that', 'staying', 'deli'], [])	DEFAULT	
(['seat', 'belt'], ['seatbelt'], ['seat', 'belt'])	Matching Pair	
(['an'], [], ['an'])	Matching Pair	
(['1st'], ['first'], ['first'])	Matching Pair	
(['7773', 'hundred'], ['triple', '7', '300'], ['triple', '7', '300'])	Matching Pair	
(['of'], ['of'], ['for'])	Matching Pair	
(['runway'], ['one', 'way'], ['one-way'])	DEFAULT	
(['have'], ['have'], ['have'])	Matching Pair	
(['hotspot'], ['hotspot'], ['hot', 'spot'])	Matching Pair	
(['we'], ['we'], [])	Matching Pair	
(['notice'], ['noticed'], ['noticed'])	Matching Pair	
(['a380'], ['a', '380'], ['a380'])	Matching Pair	
(['fuse', 'large'], ['fuse', 'large'], ['fuselage'])	Matching Pair	
(['spacy'], ['spacey'], ['spacey'])	Matching Pair	
(['pyjamas'], ['pajamas'], ['pajamas'])	Matching Pair	
([], ['yeah'], [])	Matching Pair	
(['feet'], [], ['feet'])	Matching Pair	
(['shields', 'to', 'what's', 'your', 'seat'], ['just', 'to', 'what', 'you're', 'if', 'you', 'can't', 'reach', 'the', 'seat', 'in', 'front', 'of', 'you'], ['heels', 'towards', 'your', 'seat', 'if', 'you', 'can't', 'reach', 'the', 'seat', 'in', 'front', 'of', 'you'])	DEFAULT	

([], ['and', 'lean'], ['and', 'lean'])	Matching Pair	
([], ['place'], ['place'])	Matching Pair	
([], ['in', 'an'], ['in', 'an'])	Matching Pair	
(['lead'], ['get'], ['lead'])	Matching Pair	
(['nearest', 'exit'], ['nerves', 'no', 'mm', 'okay', 'yeah', 'yeah', 'yeah', 'yeah', 'good', 'morning'], ['nearest', 'exit', 'do', 'so'])	DEFAULT	
(['8'], ['eight'], ['eight'])	Matching Pair	
(['cubbies'], ['cubbies'], ['copies'])	Matching Pair	
(['seatback'], ['seat', 'back'], ['seat', 'back'])	Matching Pair	
([], ['uh', 'huh', 'yeah', 'yeah'], [])	Matching Pair	
(['inflight'], ['in', 'flight'], ['in-flight'])	DEFAULT	
(['2577'], ['2,577'], ['2577'])	Matching Pair	
(['five'], ['five'], ['5'])	Matching Pair	
(['at'], [], [])	Matching Pair	
([], ['okay', 'yeah', 'okay', 'okay', 'okay', 'yeah', 'yeah', 'yeah', 'okay', 'yeah', 'okay', 'yeah'], [])	Matching Pair	
([], ['yeah', 'no', 'yeah', 'good', 'job'], [])	Matching Pair	
(['9'], ['nine'], ['nine'])	Matching Pair	
(['10'], ['10'], ['ten'])	Matching Pair	
([], ['okay', 'okay', 'yeah', 'okay', 'okay'], [])	Matching Pair	
(['nightcap'], ['night', 'cap'], ['nightcap'])	Matching Pair	
([], ['yeah'], [])	Matching Pair	
(['in', 'flight'], ['in', 'flight'], ['in-flight'])	Matching Pair	
(['an'], ['and'], ['and'])	Matching Pair	
(['3'], ['three'], ['three'])	Matching Pair	
(['changi'], ['shanghai'], ['changi'])	Matching Pair	
(['strait'], ['strait'], ['straits'])	Matching Pair	
([], ['okay', 'yeah', 'yeah', 'yeah', 'yeah', 'yeah', 'yeah', 'yeah', 'mm', 'okay'], [])	Matching Pair	
(['a'], ['eight'], ['a'])	Matching Pair	
(['red', 'eye'], ['red', 'eye'], ['red-eye'])	Matching Pair	
(['i'], ['i've'], ['i'])	Matching Pair	
(['and'], ['to'], ['to'])	Matching Pair	
(['i've'], ['i'], ['i've'])	Matching Pair	
(['a380'], ['a', '380'], ['a380'])	Matching Pair	

(['awhile'], ['a', 'while'], ['a', 'while'])	Matching Pair	
(['how'], ['helped'], ['how'])	Matching Pair	
(['rest'], ['rest'], [])	Matching Pair	
(['3', 'full'], ['34'], ['three', 'full'])	FURTHER ALIGN	['three', 'full']
(['surfshark'], ['serve', 'shark'], ['surf', 'shark'])	FURTHER ALIGN	['surf', 'shark']
(['forward'], ['four'], ['four'])	Matching Pair	
([], [], ['percent'])	Matching Pair	
(['surfshark', 'vpn'], ['serve', 'shark', 'vpn'], ['surfsharkvpn'])	FURTHER ALIGN	['serve', 'shark', 'vpn']

v4:

value (Azure, AWS, YouTube)	Condition	Aligned Value
(['to'], ['two'], ['to'])	Matching Pair	
(['a'], ['a'], ['in'])	Matching Pair	
(['shipping'], ['chipping'], ['chipping'])	Matching Pair	
(['60', 'degree'], ['60°'], ['60-degree'])	DEFAULT	
(['edge', 'is'], ['edges'], ['edge', 'is'])	Matching Pair	
(['ahead'], ['ahead'], ['a', 'head'])	Matching Pair	
(['and'], ['and'], ['it', "doesn't"])	Matching Pair	
([], ['it'], [])	Matching Pair	
(['are', 'are'], ['are', 'are'], ['are'])	Matching Pair	
([], ['and'], ['and'])	Matching Pair	
(['lower'], ['lower'], ['low', 'or'])	Matching Pair	
(['shots', 'are'], ['shot', "they're"], ['shot', "they're"])	Matching Pair	
(['cause'], ['because'], ['because'])	Matching Pair	
(['ship'], ['chip'], ['chip'])	Matching Pair	
(['yet'], ['you', 'have'], ['yet'])	Matching Pair	
(['volume'], ['the', 'ball', 'in'], ['the', 'ball', 'in'])	Matching Pair	

(['their', 'your', 'the'])	SINGLE WORD	
(['we', 'it', 'we'])	Matching Pair	
(['high', 'how', 'high'])	Matching Pair	
(['hole', 'whole', 'hole'])	Matching Pair	
(['going', 'to', 'gonna', 'gonna'])	Matching Pair	
(['hi', 'high', 'a', 'hi', 'a'])	DEFAULT	
(['when', 'you', 'chip', 'you', ['when', 'you', 'chip', 'you'], ['in', 'each', 'if', 'you']])	Matching Pair	
(['front', 'foot', 'front', 'foot'], ['front', 'foot', 'front', 'foot'], ['front', 'foot', 'front', 'foot'])	Matching Pair	
(['ahead', 'ahead', 'a', 'head'])	Matching Pair	
(["you'll"], ['you', 'will', 'you', 'will'])	Matching Pair	
(['in', 'and', 'in'])	Matching Pair	
(['well', 'while', 'well'])	Matching Pair	