



# IIT Madras

ONLINE DEGREE

**Computational Thinking**  
**Professor. Madhavan Mukund**  
**Department of Computer Science**  
**Department of Mathematical Institute, Chennai**  
**Professor G. Venkatesh**  
**Indian Institute of Technology, Madras**

**Lecture 3.3**

**Systematic process of hypothesis verification to find the relation between word length and frequency**

Professor. Madhavan Mukund: So, we had looked at these word frequencies earlier and we had found out for instance that these were the words which are there in this paragraph which have more than 1 occurrence and every other words in this paragraph has only 1 occurrence. So, one thing we have noticed looking at these words is that they seem to be. Except for Monday which has 6 letters, they all seem to be small words. So, do you think it is the case that the frequent words are always the small words?

Professor. G. Venkatesh: They are likely no, looks like it because longer words we will tend not to barely see, we when for example, Vedanayagam it is a very long word.

Professor. Madhavan Mukund: Yes.

Professor. G. Venkatesh: And you tend to say his his his. Way to write Vedanayagam but with out writing Vedanayagam. So, it is slightly that the words that are occurring more than once very often are shorter and ones which are longer occur less frequently or one ones maybe. This case only ones.

Professor. Madhavan Mukund: So, frequently we decided was at least twice, so we know how to talk about frequent and not frequent.

So, frequent is anything which is on this list more than once. All the other words are infrequent and now we want to ask about shorter long words. So, short.

Professor. Madhavan Mukund: So, below means.

Professor. Madhavan Mukund: So, then, how should we find the average?

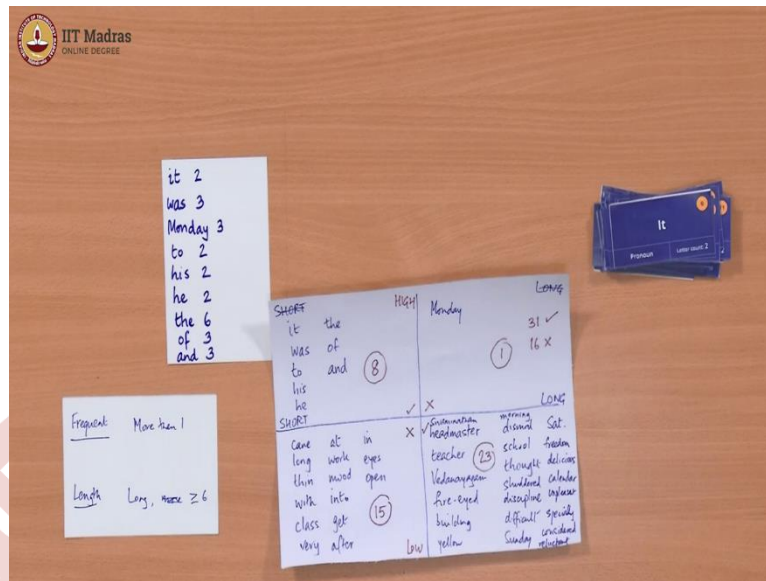
Professor. Madhavan Mukund: So, each word we know the length because letter count is there.

Professor. G. Venkatesh: So, we will take that and find the average.

Professor. Madhavan Mukund: So, meaning we have to find...

Professor. G. Venkatesh: So, add up the letter counts.

Professor. Madhavan Mukund: Add up the letter counts and divide by the total number of words. So, total number of words cards we know because they are numbers sequentially. 60 something. 0 to 64, so they are 65 cards. So, if we add up all the letters and divide by 65, we



should get an estimate of the average word length. Then we move anything smaller than that is the short one, everything bigger than that is a long one. Then we will see.

Professor. G. Venkatesh: So, let us do this addition is going to be a little.

Professor. Madhavan Mukund: So, we will just this is standard iteration, we will just go through. So, I will just keep reading out the letters, you keep adding it. So, I start with 2.

Professor. G. Venkatesh: 2, 3.

Professor. Madhavan Mukund: 5.

Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 11.

Professor. G. Venkatesh: 7

Professor. Madhavan Mukund: 18.

Professor. G. Venkatesh: 11.

Professor. Madhavan Mukund: 29.

Professor. G. Venkatesh: 3

Professor. Madhavan Mukund: 32.

Professor. G. Venkatesh: 9

Professor. Madhavan Mukund: 41.

Professor. G. Venkatesh: 9.

Professor. Madhavan Mukund: 43.

Professor. G. Venkatesh: 4

Professor. Madhavan Mukund: 47.

Professor. G. Venkatesh: 3

Professor. Madhavan Mukund: 50.

Professor. G. Venkatesh: 4

Professor. Madhavan Mukund: 54.



Professor. G. Venkatesh: 2

Professor. Madhavan Mukund: 56.

Professor. G. Venkatesh: 10

Professor. Madhavan Mukund: 66.

Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 72.

Professor. G. Venkatesh: 9.

Professor. Madhavan Mukund: 81.

Professor. G. Venkatesh: 10.

Professor. Madhavan Mukund: 91.

Professor. G. Venkatesh: 2.

Professor. Madhavan Mukund: 93.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 96.

Professor. G. Venkatesh: 8.

Professor. Madhavan Mukund: 104.

Professor. G. Venkatesh: 5.

Professor. Madhavan Mukund: 109.

Professor. G. Venkatesh: 3.

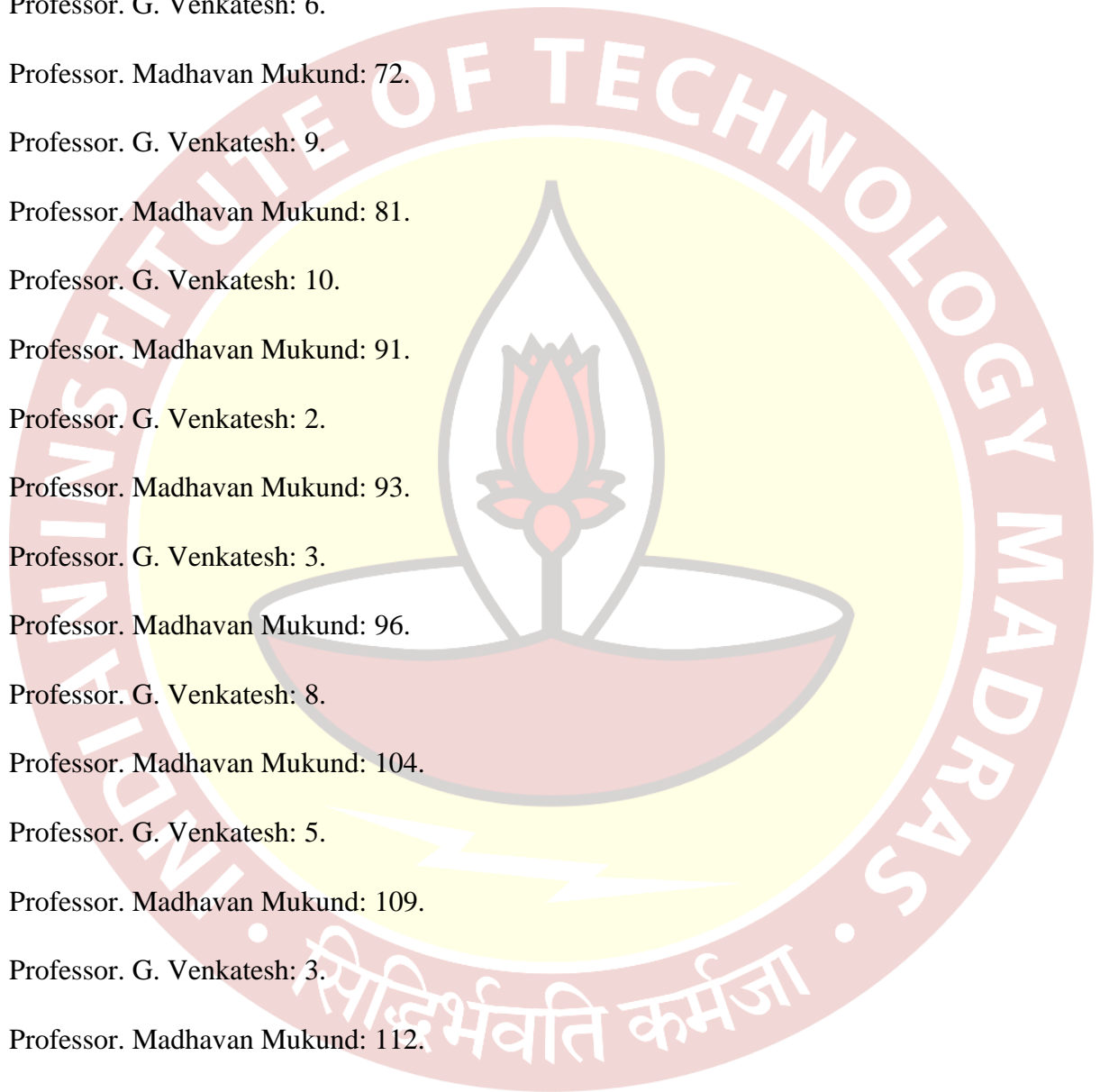
Professor. Madhavan Mukund: 112.

Professor. G. Venkatesh: 9.

Professor. Madhavan Mukund: 121.

Professor. G. Venkatesh: 7.

Professor. Madhavan Mukund: 128.



Professor. G. Venkatesh: 2

Professor. Madhavan Mukund: 130.

Professor. G. Venkatesh: 8.

Professor. Madhavan Mukund: 138.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 141.

Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 147.

Professor. G. Venkatesh: 2.

Professor. Madhavan Mukund: 149.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 152.

Professor. G. Venkatesh: 9.

Professor. Madhavan Mukund: 161.

Professor. G. Venkatesh: 2.

Professor. Madhavan Mukund: 163.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 166.

Professor. G. Venkatesh: 4.

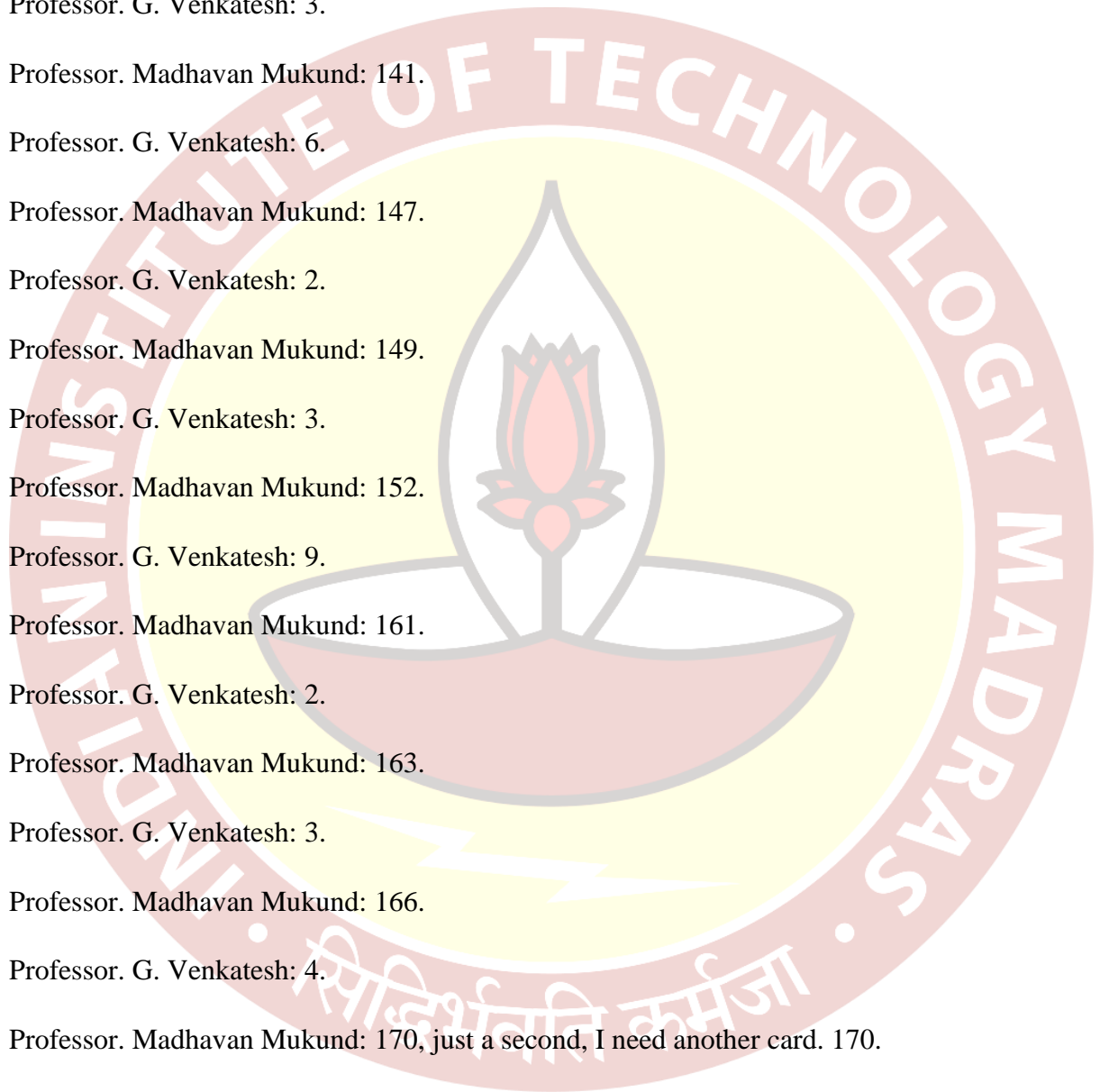
Professor. Madhavan Mukund: 170, just a second, I need another card. 170.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 173.

Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 179.



Professor. G. Venkatesh: 4.

Professor. Madhavan Mukund: 182.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 185.

Professor. G. Venkatesh: 4.

Professor. Madhavan Mukund: 189.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 192.

Professor. G. Venkatesh: 10.

Professor. Madhavan Mukund: 202.

Professor. G. Venkatesh: 2.

Professor. Madhavan Mukund: 204.

Professor. G. Venkatesh: 9.

Professor. Madhavan Mukund: 213.

Professor. G. Venkatesh: 2.

Professor. Madhavan Mukund: 215.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 218.

Professor. G. Venkatesh: 4.

Professor. Madhavan Mukund: 222.

Professor. G. Venkatesh: 7.

Professor. Madhavan Mukund: 229.

Professor. G. Venkatesh: 2.

2Professor. Madhavan Mukund: 231.



Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 237.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 240.

Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 246.

Professor. G. Venkatesh: 6.

Professor. Madhavan Mukund: 252.

Professor. G. Venkatesh: 8.

Professor. Madhavan Mukund: 260.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 263.

Professor. G. Venkatesh: 8.

Professor. Madhavan Mukund: 271.

Professor. G. Venkatesh: 11.

Professor. Madhavan Mukund: 282.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 285.

Professor. G. Venkatesh: 5.

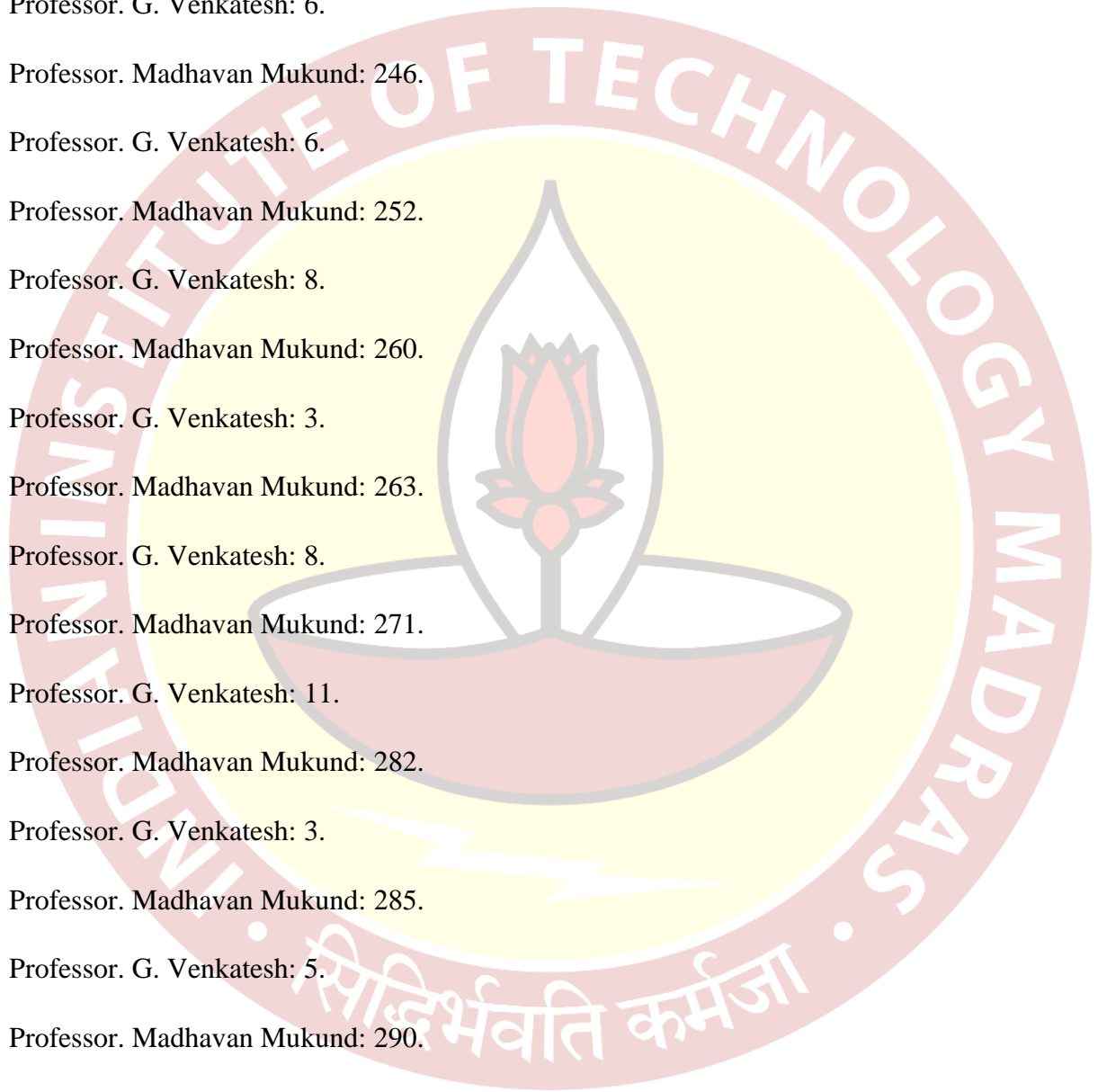
Professor. Madhavan Mukund: 290.

Professor. G. Venkatesh: 7.

Professor. Madhavan Mukund: 297.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 300.





Professor. G. Venkatesh: 10.

Professor. Madhavan Mukund: 310.

Professor. G. Venkatesh: 4.

Professor. Madhavan Mukund: 314.

Professor. G. Venkatesh: 3.

Professor. Madhavan Mukund: 317.

Professor. G. Venkatesh: 4.

Professor. Madhavan Mukund: 321.

Professor. G. Venkatesh: 4.

Professor. Madhavan Mukund: 325.

Professor. G. Venkatesh: And another 4.

Professor. Madhavan Mukund: 329. So, we have 329 and 65. So, this is roughly, if it was 5 it will be 330. So, it is little under 5, so this is some 4. So, 5 into 365 is, 5 into 65 is 325.

Professor. G. Venkatesh: This is about 5.

Professor. Madhavan Mukund: So, we can say 5 and less is short and 6 and above is long. So, now we need to count how many will be have.

Professor. G. Venkatesh: We have 2 card parameters . Short long.

Professor. Madhavan Mukund: So, we have now frequent, so we will write it down. We have frequent and we have length. So, frequent we said it is more than 1, more than 1 is frequent and long is greater than more than 6, greater than or equal to 6.

Professor. G. Venkatesh: So, greater than or equal to 6.

Professor. Madhavan Mukund: So now we want to look at these combinations, which words are frequent and which words are long.

Professor. G. Venkatesh: So, there are 4 possibilities.

Professor. Madhavan Mukund: So what we should do is, so we should probably think of this in terms of maybe creating so, on this direction we have say short and long.

Professor. G. Venkatesh: So, like this short one here and long over here.

Professor. Madhavan Mukund: So, left is short, right is long. And in this direction maybe we use a different color. So, in this direction we have maybe I should written it in the centre, that is probably better. So, high and low and let us write it here so that is clear, short and long. So, these are our criteria.

Professor. G. Venkatesh: So, basically high frequency words are going to be written on the top half and low frequency words bottom half and long words on the right and short words on the left. So, we keep writing. I read out the word, you write or just the sequence number of card is enough or write word.

Professor. Madhavan Mukund: I think we can write, just write 1 something.

Professor. G. Venkatesh: We need only to keep count.

Professor. Madhavan Mukund: So, in a sense we have these words which we already know their frequency. We know that these are all the high frequency words. Everything else low frequency.

Professor. G. Venkatesh: So, they are all on this side.

Professor. Madhavan Mukund: So, they are all on this side. So, we can count and put them on short and long. So, let us write the words. So, we have here we have it because it is a short word because it high frequency.

Professor. G. Venkatesh: So, it is high frequency short word.

Professor. Madhavan Mukund: Was. Now Monday is 6 letters, we said 6 letters is a long word. So, it is high frequency but it is a long word. So, Monday goes here. Then we have to, his, he, the, of, and. So, among these words,

Professor. G. Venkatesh: So, among the high frequency words, so how many were there, totally they were 1, 2, 3, 4, 5, 6, 7, 8, 9 words.

Professor. Madhavan Mukund: 9 words.

Professor. G. Venkatesh: Out of 9, 8 of them happen to be short and one so exception is this 8 of conforming our.

Professor. Madhavan Mukund: So, what we said is it is more likely that high frequency word is short, one is one but this one is refuting it is not conforming, it is refuting our hypothesis.

So, now we should see where you have to do the other side, where the remaining words come.

Professor. G. Venkatesh: If lot of those words also come here then we are in trouble.

Professor. Madhavan Mukund: So, we know that any word which is not here is a low this should come below the line which is then we decide left or right.

Professor. G. Venkatesh: So, we just have to count the number of letters and write it in this or this. Cane.

Professor. Madhavan Mukund: So, cane is left.

Professor. G. Venkatesh: long.

Professor. Madhavan Mukund: long is left.

Professor. G. Venkatesh: thin.

Professor. Madhavan Mukund: thin is left.

Professor. G. Venkatesh: his.

Professor. Madhavan Mukund: his is already done. It is already high frequency.

Professor. G. Venkatesh: with.

Professor. Madhavan Mukund: with is here.

Professor. G. Venkatesh: headmaster.

Professor. Madhavan Mukund: Headmaster is here.

Professor. G. Venkatesh: And.

Professor. Madhavan Mukund: And is done.

Professor. G. Venkatesh: Teacher.

Professor. Madhavan Mukund: Teacher is here.

Professor. G. Venkatesh: Class.

Professor. Madhavan Mukund: Class is 5 letters so it is.

Professor. G. Venkatesh: His done. Vedanayagam

Professor. Madhavan Mukund: okay.

Professor. G. Venkatesh: Fire-eyed, the, building.

Professor. Madhavan Mukund: Building.

Professor. G. Venkatesh: Yellow.

Professor. Madhavan Mukund: Yellow.

Professor. G. Venkatesh: Dismal.

Professor. Madhavan Mukund: Dismal is 6 letters.

Professor. G. Venkatesh: The is done, school.

Professor. G. Venkatesh: School is also 6 letters.

Professor. G. Venkatesh: Of.

Professor. Madhavan Mukund: Of is done.

Professor. G. Venkatesh: Thought.

Professor. Madhavan Mukund: Thought is a long word.

Professor. G. Venkatesh: Very.

Professor. Madhavan Mukund: Short word.

Professor. G. Venkatesh: we do have short words there, the is done. At.

Professor. Madhavan Mukund: At is also short. Its strange that at is only 1

Professor. G. Venkatesh: Shuddered.

Professor. Madhavan Mukund: Shuddered.

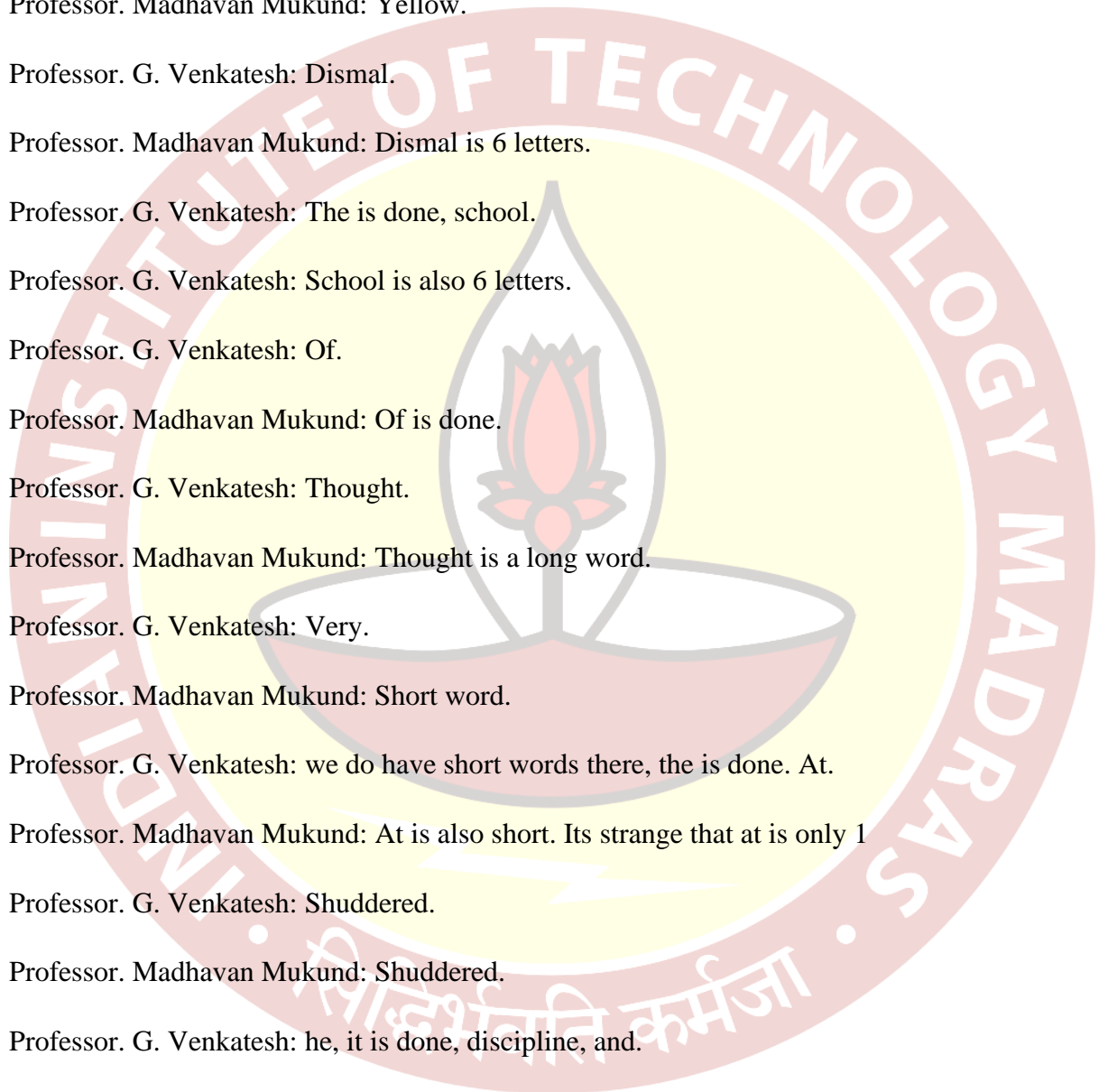
Professor. G. Venkatesh: he, it is done, discipline, and.

Professor. Madhavan Mukund: And is done.

Professor. G. Venkatesh: Work.

Professor. Madhavan Mukund: Work is here.

Professor. G. Venkatesh: of is done.



Professor. Madhavan Mukund: Of is done.

Professor. G. Venkatesh: Mood.

Professor. Madhavan Mukund: Mood is here.

Professor. G. Venkatesh: Monday.

Professor. Madhavan Mukund: Monday is done.

Professor. G. Venkatesh: The,

Professor. Madhavan Mukund: The is done.

Professor. G. Venkatesh: Into.

Professor. Madhavan Mukund: Into is only ones. Ok, strange

Professor. G. Venkatesh: Get.

Professor. Madhavan Mukund: Get.

Professor. G. Venkatesh: To.

Professor. Madhavan Mukund: To is done.

Professor. G. Venkatesh: Difficult.

Professor. Madhavan Mukund: Difficult is here.

Professor. G. Venkatesh: Was.

Professor. Madhavan Mukund: Was is done.

Professor. G. Venkatesh: It.

Professor. Madhavan Mukund: It is done.

Professor. G. Venkatesh: Sunday.

Professor. Madhavan Mukund: Sunday is here.

Professor. G. Venkatesh: And.

Professor. Madhavan Mukund: Done. And is done.

Professor. G. Venkatesh: Saturday.

Professor. Madhavan Mukund: I will write it as short. Saturday.

Professor. G. Venkatesh: Of.

Professor. Madhavan Mukund: Of is done.

Professor. G. Venkatesh: Freedom.

Professor. Madhavan Mukund: Freedom.

Professor. G. Venkatesh: Delicious.

Professor. Madhavan Mukund: Delicious.

Professor. G. Venkatesh: The is done.

Professor. Madhavan Mukund: Yes.

Professor. G. Venkatesh: After, after is short. Calendar, the.

Professor. Madhavan Mukund: Done.

Professor. G. Venkatesh: In is..

Professor. Madhavan Mukund: In is only one.

Professor. G. Venkatesh: Surprising. Unpleasant, specially.

Professor. Madhavan Mukund: Specially.

Professor. G. Venkatesh: Monday is done. Considered.

Professor. Madhavan Mukund: Considered.

Professor. G. Venkatesh: He.

Professor. Madhavan Mukund: He is done.

Professor. G. Venkatesh: Eyes.

Professor. Madhavan Mukund: Eyes is short.

Professor. G. Venkatesh: His is done. Open.

Professor. Madhavan Mukund: Open is short.

Professor. G. Venkatesh: To.

Professor. Madhavan Mukund: To is done.

Professor. G. Venkatesh: Reluctant.

Professor. Madhavan Mukund: Reluctant I have to write it somewhere, let me write it here, it is kind of bit messy but I...

Professor. G. Venkatesh: Was.

Professor. Madhavan Mukund: Was is done.

Professor. G. Venkatesh: Swaminathan.

Professor. Madhavan Mukund: The most important one, Swaminathan.

Professor. G. Venkatesh: Morning.

Professor. Madhavan Mukund: Morning is here.

Professor. G. Venkatesh: Monday is done. Was is done.

Professor. Madhavan Mukund: Was is done.

Professor. G. Venkatesh: It is done.

Professor. Madhavan Mukund: So, let us count.

Professor. G. Venkatesh: So, this is not conformed, not conforming because it is a frequent word.

Professor. Madhavan Mukund: Frequent word.

Professor. G. Venkatesh: but long.

Professor. Madhavan Mukund: But long.

Professor. G. Venkatesh: What about these, what can we say about these? These are infrequent words.

Professor. Madhavan Mukund: Which are long. So, this is also conforming.

Professor. G. Venkatesh: Conformed.

Professor. Madhavan Mukund: Because we are saying..

Professor. G. Venkatesh: Short words are frequent, long words are infrequent.



Professor. Madhavan Mukund: Yes.

Professor. G. Venkatesh: So, these are the again conforming.

Professor. Madhavan Mukund: Yes.

Professor. G. Venkatesh: These words are short words but infrequent.

Professor. Madhavan Mukund: So, they do not conform.

Professor. G. Venkatesh: They do not conform.

Professor. Madhavan Mukund: So, they are refuting the hypothesis.

Professor. G. Venkatesh: So, in some sense these and this are refuting and this and this are conforming.

Professor. Madhavan Mukund: So, if we just picture it, so this is again star hypothesis.

Professor. G. Venkatesh: This is again star hypothesis.

Professor. Madhavan Mukund: This is also again star hypothesis.

Professor. G. Venkatesh: Again.

Professor. Madhavan Mukund: And these two boxes I mean are for hypothesis.

Professor. G. Venkatesh: So, we should add this, add this, let us put the number in the boxes, maybe that is....

Professor. Madhavan Mukund: So, we have 1, 2, 3, 4, 5, 6, so we have 8 words over here. We have 1 word. So, let us circle it. So, we have 1 word over here. Here we have, 1, 2, 3, 6, 9, 10, 11, 12, 13, 14, 15 over here and this is more necessary. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23. We have 23 over here. So, totally we have 9, 24, 47 distinct words in those 65. And what we are saying is that 31 confirm, 31 are good words and 16 are refuting. So, almost two third are conforming.

Professor. G. Venkatesh: So, more we can say more or less. So, it is more or less. Two thirds are conforming an hypothesis and in fact actually if you look at it from one direction, which means that the high frequency words tend to be short, high frequency words tend to be short.

Professor. Madhavan Mukund: Tend to be short.

Professor. G. Venkatesh: That then it conforms.



Professor. Madhavan Mukund: Then it is very strong because 8 put of 9 high frequency words are there.

Professor. G. Venkatesh: So, we seem to have 90 percent conforming that hypothesis.

Professor. Madhavan Mukund: Yes.

Professor. G. Venkatesh: Other hand if you say low frequency words are long, then 23 by 48 so that is not too much. It is about 38.

Professor. Madhavan Mukund: 23 by 38. So, it is roughly 60 percent again.

Professor. G. Venkatesh: 60 percent. So, it is kind of more than 50 percent, 60 percent, so it is a kind of okay but not a very strong conforming like this one. But overall if you look two thirds of the things are conforming the hypothesis that there is a correlation between that length of the word and the frequency.

Professor. Madhavan Mukund: Frequency. Interesting.

