- Does sentence delimitations mean dot ? If yes I have done the task. So why I do not pass yet?

Your manipulations are not as much important as your reflection on the sentence delimitations.

Sentence boundaries are defined in the files by the extra empty line you add after whatever sign you think is the sentence limit (dots, but also some other punctuations).

Anyway, even if your file confirm that the manipulations are done, I still need you to show me a reflection on sentence boundary disambiguation in your file VS in the gold standard. Do you know for instance if, in the gold standard, there have been settled more sentence delimitations than in yours? Do you know why or why not? (tips: take just one or two examples of sentence delimitation difference in your file and in facit.txt and try to explain me why your sentence disambiguation program did not manage to do the same).

- I have no scanner. How can I give you back the transducer exercise?

You can perfectly make your transducer by hand on a simple paper no need for a scanner or a printer. Make sure I get it by droping it in my mailbox on the third floor. (and keep a draft, so you can redo it easily in case you fail on this part). If you still get a problem for transmiting me the work please tell me before the deadline, we always find a solution !

- When is corrected my work?

I starts the correction after the deadline. The speed depends each years on the workflow it takes between 2 and 3 weeks generally.

- If I fail what happen?

I explain you why you have failed. You get to correct/redo only the part you have failed. Sometimes students fail because they just forget to give back a part of the assignment (sentence segmentation and transducer mostly). Please give it a try!

- I have a questions about the lexical forms . If a word is a Swedish word, so can I really use the English way of putting tags on this word? Or is it ok to use SUC (the swedish tagset)?

Yes you can use SUC tag set. As we said in class, tag sets are not universal, they are language specific. If you use the standard morphosyntactic feature set used on the SUC corpus, don't hesitate to say it in your report. Or, if you use another standard, make precise briefly what your features mean (for instance, "IND"=indicative). It is indeed not standardized so we have to make it precise.

- For the transducer exercise I have the word “blabla” and I have to add/remove a “a” . Should I write my transducer like that: “blabla1” or like that “blabla2” ?

Are you sure your question is about a notion? Your teacher cannot solve the exercise for you. I can give complete feedback about your specific answer only when I correct the reports.

- I have two words to transduce. Should I make one or two transducers?

I ask for one transducer that can handle both. Yes it is possible. We saw in class examples of transducers that could handle more than just one word.

- Should my transducer be deterministic?

No. Sometimes it is not even possible to make it deterministic.

- I have “12345” tokens/sentences in my file and there is “5678”(note:dummy numbers) tokens/sentences in facit.txt . Is this ok? I have still words like “blabla--blabla” that are wrongly tokenized, should I handle them?

The answer is not so simple.
Recall that Facit.txt is humanly done thus it takes into account every linguistics exceptions. Don't forget that tokenization is an endless task, thus you want to be right most of the time but you know that you will never have the perfect tokenizer, you decide when you should stop to take into account the exceptions. For the correction of the assignment I hope to see something better than just separation based on spaces, but I definitely don't hope you to make me the 95% accuracy tokenizer/sentence splitter.