

Summary / Synopsis Guide (DECODA corpus inspired version)

Why do we need this document?

Making a summary is a very subjective task. Two different persons won't systematically pick up the same information, or the same level of detail, and then won't write the same summary.

Our systems need these human summaries to learn how to automatically generate them in a good way, and to be evaluated as well. Here we have two options:

- The first option is to hire a lot of annotators and let them write their summaries in a free way. And then make a study on all the resources collected to find a good shape for a summary. But this method is very expensive.
- The second option is to write a guide to lead the annotators in a certain way of annotation to get some similarity in the summaries generated. In this way we introduce some guidelines depending on the corpus to be annotated (previously studied). The annotators are not free anymore, but then only a few can annotate the corpus. Because of that kind of guide, all the summaries generated will come with some similarity and become more usable for our systems.

DECODA Corpus inspired version.

First of all, let's define here what we call a synopsis. A synopsis is a summary of the call taken from a call center. This call mostly involved two persons, the caller (here someone who wants something from the call center) and the adviser (the one who answers in the call center and gives some solution to the caller).

After some studies we concluded that this synopsis should be no longer than 7% of the original call (in terms of words). We are aware that the annotator won't spend their time counting the words... So we probably include this length limit in the next version of the interface. Speaking of the interface, right now, it's just a simple and minimalist interface that just provides the conversation (spoken and written) and a box to fill with the synopsis. (Everything should be revamped to a better version).

Go back to the synopsis, this is a pretty subjective thing to do, that's why we'll try in this guide to give some lines to follow.

First we can distinguish two kinds of synopsis:

- The semantic oriented synopsis: that focused more on the content of the conversation than the end of it.
- The "structure" oriented synopsis: That focused more on the way the adviser treats the caller and his call.

For the rest of this guide, only the semantic oriented synopsis we'll be considered (the "structure" oriented synopsis could be developed in an other guide).

Syntax, length, way to write.

You have to keep in mind that you are limited in term of length. Moreover we don't really need to get some really good language, what we need the most is the information!

Here's some tips:

- Only pick the important information in the call.
- The shorter your sentence is the better your synopsis will be
- Do not hesitate to make a "telegraphic style" sentence (e.g. "Route request in Paris center")
- Try to sum the long exchange with a simple action, or if there isn't any good information in it don't even pick it up in the summary.

Practical cases.

Nothing's better than a good example. Here's three of them translated from french with some comments on how we ended here:

Example 1

- Hello
- Hello
- Hello
- eh it's Mrs [name removed] eh I was calling you because I lost my scarf, where hmm in the bus hmm 140, when
- Yes
- I left at Colombes [name] yesterday night between
- Yes you need to call later madam, around 11am, it's too early, it's not open yet.
- At 11
- 11am yes, ok?
- Ok
- See you later.

Synopsis 1:

Needs information about a scarf lost in the bus 140. Wait the service opening and call later.

Synopsis 2:

/ Maybe there is an error made by the annotator here about the bag /
Bag lost in the bus 140, but call at 11am when the service will be open.

Comments :

Here's the call is pretty clear, the caller just want to know if there is any news about her loss. Then both annotators ended with the result of the call (e.g. call later when the service is opened). We could argue a bit on it due to the length of the synopsis, but it's a pretty important information and the length of the original conversation is pretty short...

Example 2

- Hello
- please
- hello?
- Yes, hello
- hello
- I call you because I would like to have some information, when I'm at the Javel's station
- yes
- of the RER [*the name of the train*], is there a bus that could drive me closer to the Vauthier's street at Boulogne [name of the city] without taking the metro
- Vauthier's street at Boulogne-Billancourt
- yes
- just a moment, I'm looking for it
- thanks
- Madam
- yes
- you have the 72, you cross the Mirabeau's bridge to the Mirabeau's stop in way to Saint-Cloud's park.
- yes
- and you stop at the Reine Jean-Jaurès'road's stop
- Wait, at the end of the Mirabeau's bridge I take the 72
- 72
- to Saint-Cloud and you stop at the Reine Jean-Jaurès'road's stop
- Ok
- Ok thank you
- Good bye
- Good bye
- thank you

Synopsis 1:

Needs for a connection by bus between the exit of the RER and a street in Boulogne.

Synopsis 2:

Needs for a connection by bus between the RER and the Vauthier's street at Boulogne.

Synopsis 3:

Information on a potential bus connecting the Javel's station of the RER and a street in Boulogne.

Comments :

As you can see here, the three annotators picked up the same informations, only the syntax is slightly different. Some annotator are more accurate about the name of the street (i.e Vauthier's street / street in Boulogne).

Example 3

- Hello
- Yes hello madam, eh I'd like to have some information
- Yes
- I'd like to know if the bus center strike at Vitry [*city*] will continue tomorrow or if it was just today?
- No it's, it's just today, tomorrow everything is normal
- Ok thank you very much
- My pleasure
- goodbye
- Have a good day, goodbye

Synopsis 1:

Needs informations about the renewal of the strike. Normal traffic in prevision.

Synopsis 2:

Renewal of the bus center strike at Vitry [*city*], no tomorrow the traffic is normal.

Comments :

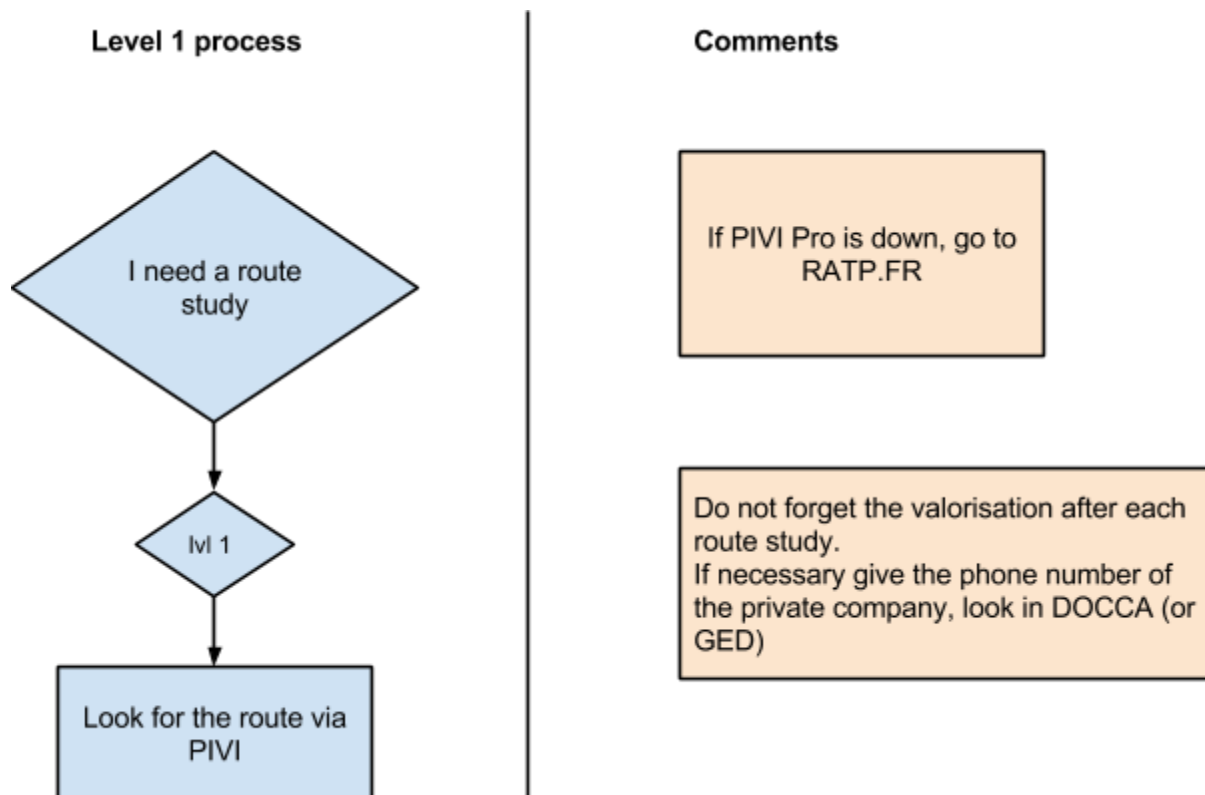
Nothing special here.

However we can notice the syntax used by annotator 1. Since the beginning he used the same method i.e. "Needs informations about". It can be a bit "word consuming" but it can also be a good way to introduce some kind of structure in every synopsis.

Call scenario

The DECODA's conversation come from a call center from the RATP. In this call center every adviser have some call scenario to help them answering the caller.

For example, the caller need to know a route, then the adviser get his route scenario and answer the question. The call scenario looks like this:



The call scenario is divided in two parts, the call process and the comments.

The comments are just some additional stuff to help the adviser during the process, it's generally very technical and then not useful for us. On the other hand the call process is very interesting.

The call process is like a state diagram. On the top on the diagram we have the main issue of the caller (here "I need a route study"). Then by following the arrow we have the way to answer depending on the other caller's issues.

Usually these call scenario draw a good base for a synopsis because if you can apply the corresponding scenario to a conversation it will give you the main theme, and the process pretty much already summed up.

How to get or generate these call scenario?

In the DECODA corpus we already got these call scenario because of the nature of the corpus. But how could we make them if they are not available for other conversations?

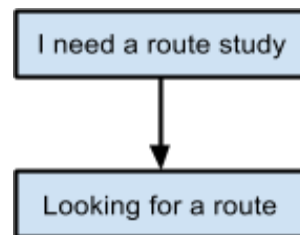
- Basically the first thing to do is to establish a list with all the main theme of the conversation.

For example in DECODA we have:

- 1 - Route
- 2 - Loses, theft, found
- 3 - Official report
- 4 - Reimbursement
- 5 - Delay, incident
- 6 - Prices
- 7 - Accidents

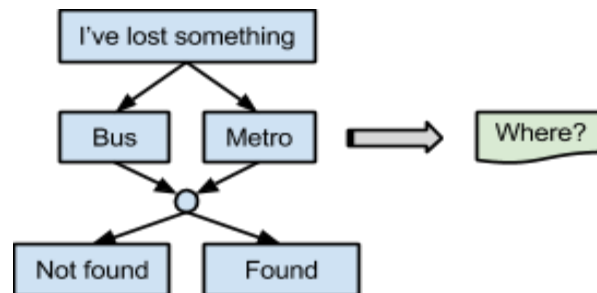
- Then for each theme you have to find if there is some redundant schema.

For instance for a route study, in most cases the caller is asking for a route and then the adviser is giving the caller the route needed. It can easily be drawn like this:



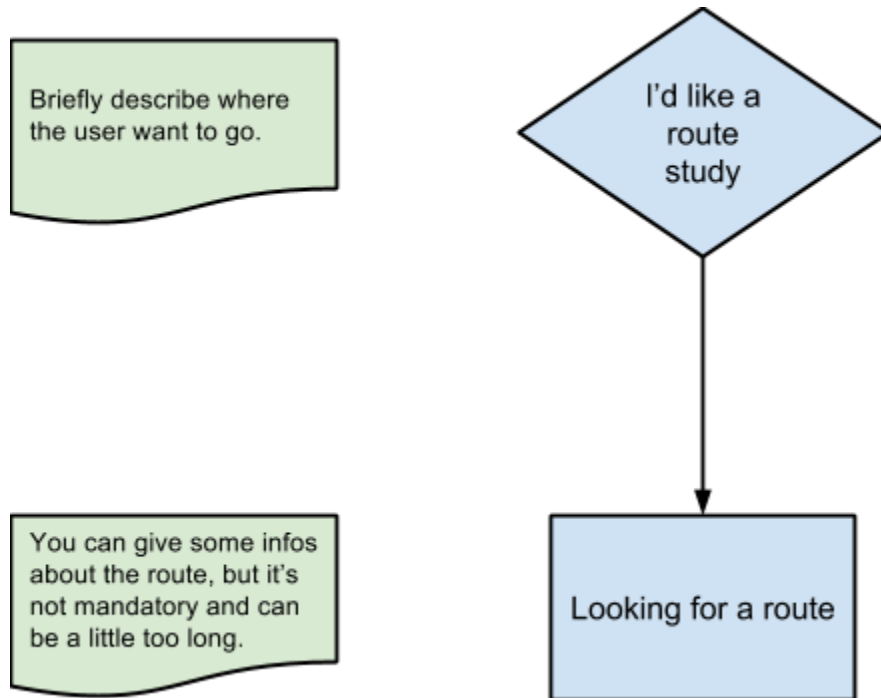
- While the state diagram can provide a lot of options, you can add some variety in it.

For instance in case of a lost, we can have several options like where the object has been lost, or is it found yet or not.



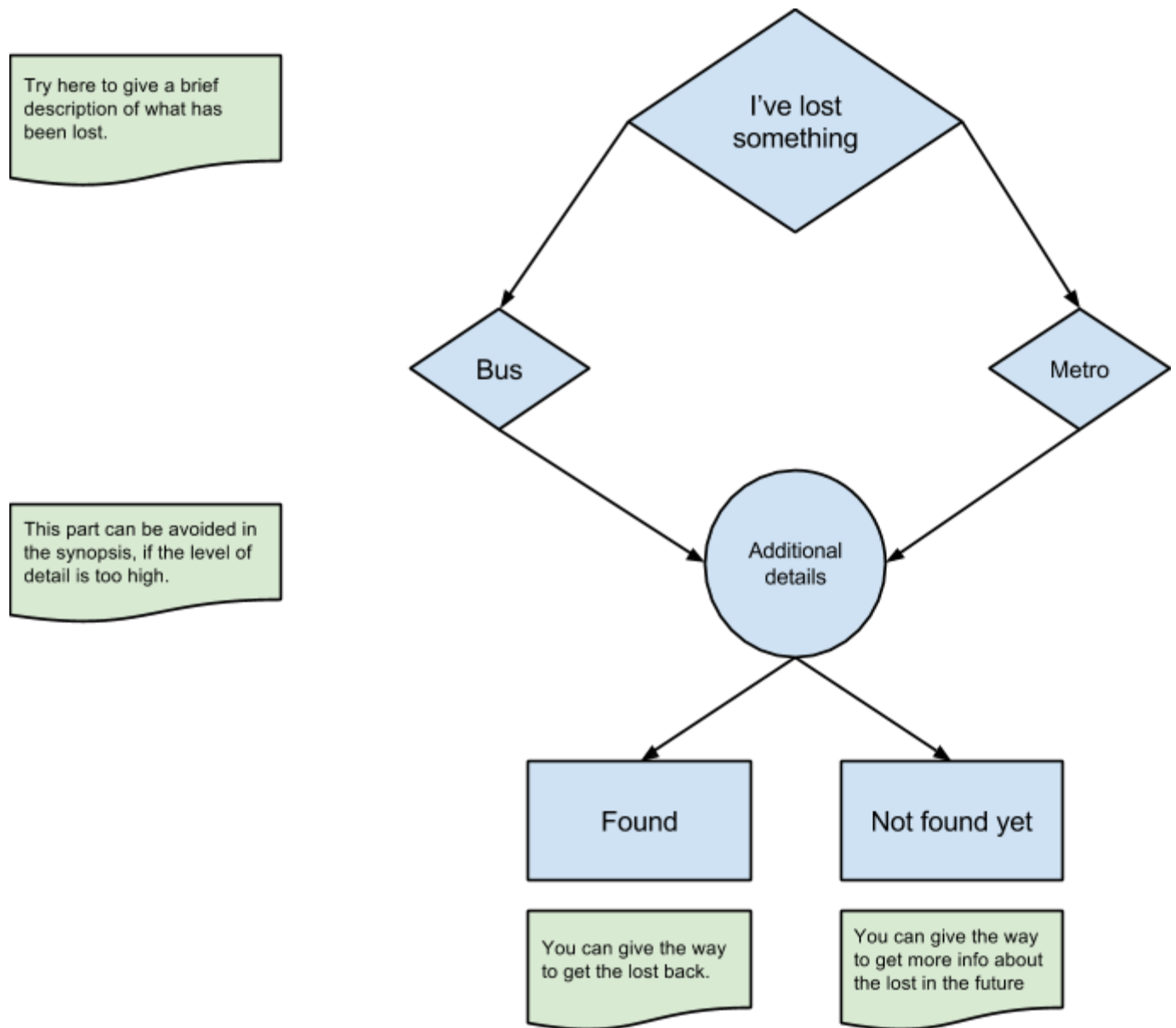
The next part gather the call scenario from DECODA listed as example above. Some comments have been added about the summarization task.

1- Route



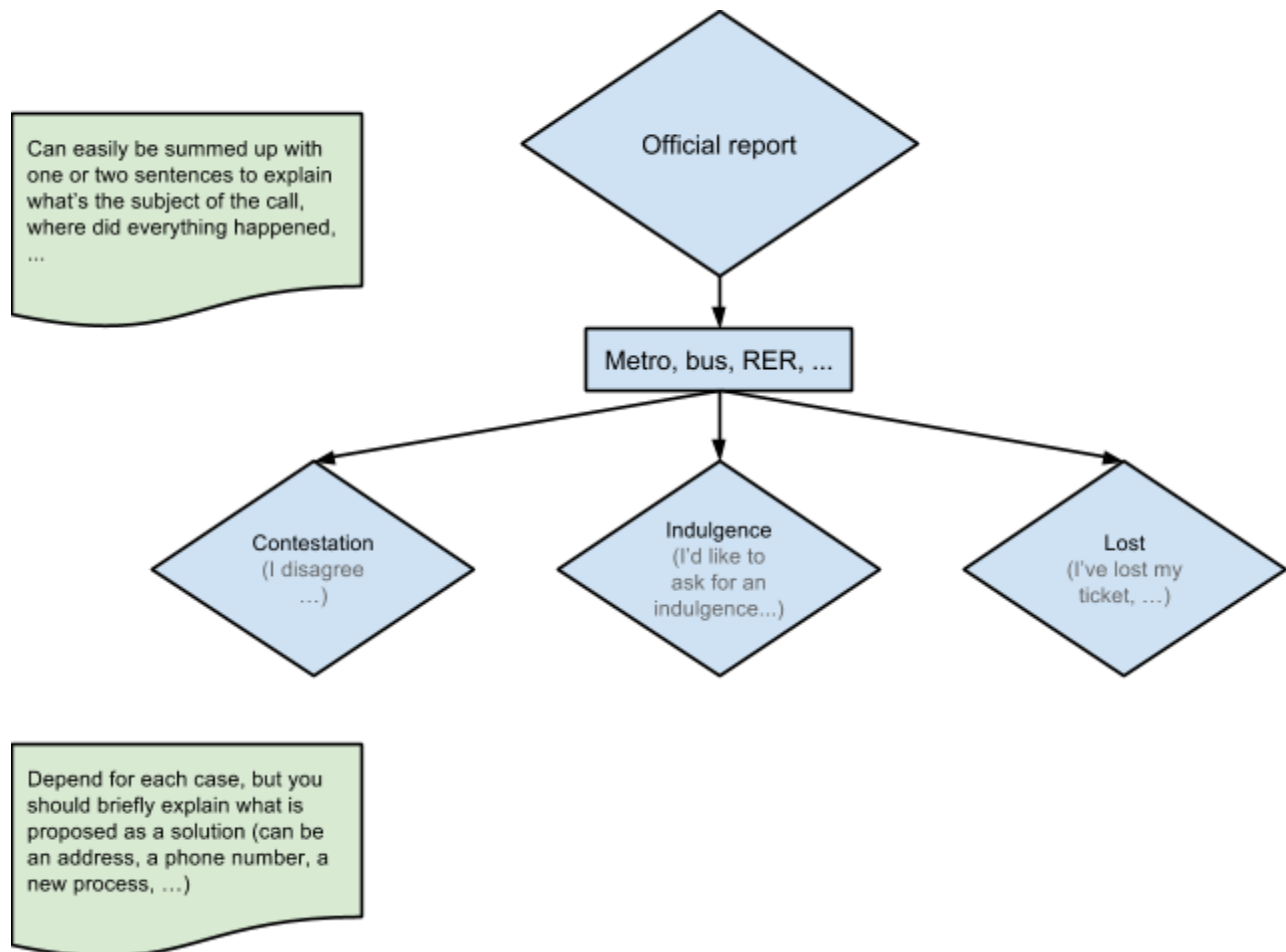
In case of several route requests in the same call, try not to focus on the destination, try to find a location where all the routes happen like for example Paris center if there are 2 or 3 route in individual different locations in Paris center.

2 - Loses, theft, found



In this particular case we like to precise if the object has been found or not just by adding at the end "object found, go to [location] to get it back" or "object not found still [recall later]".

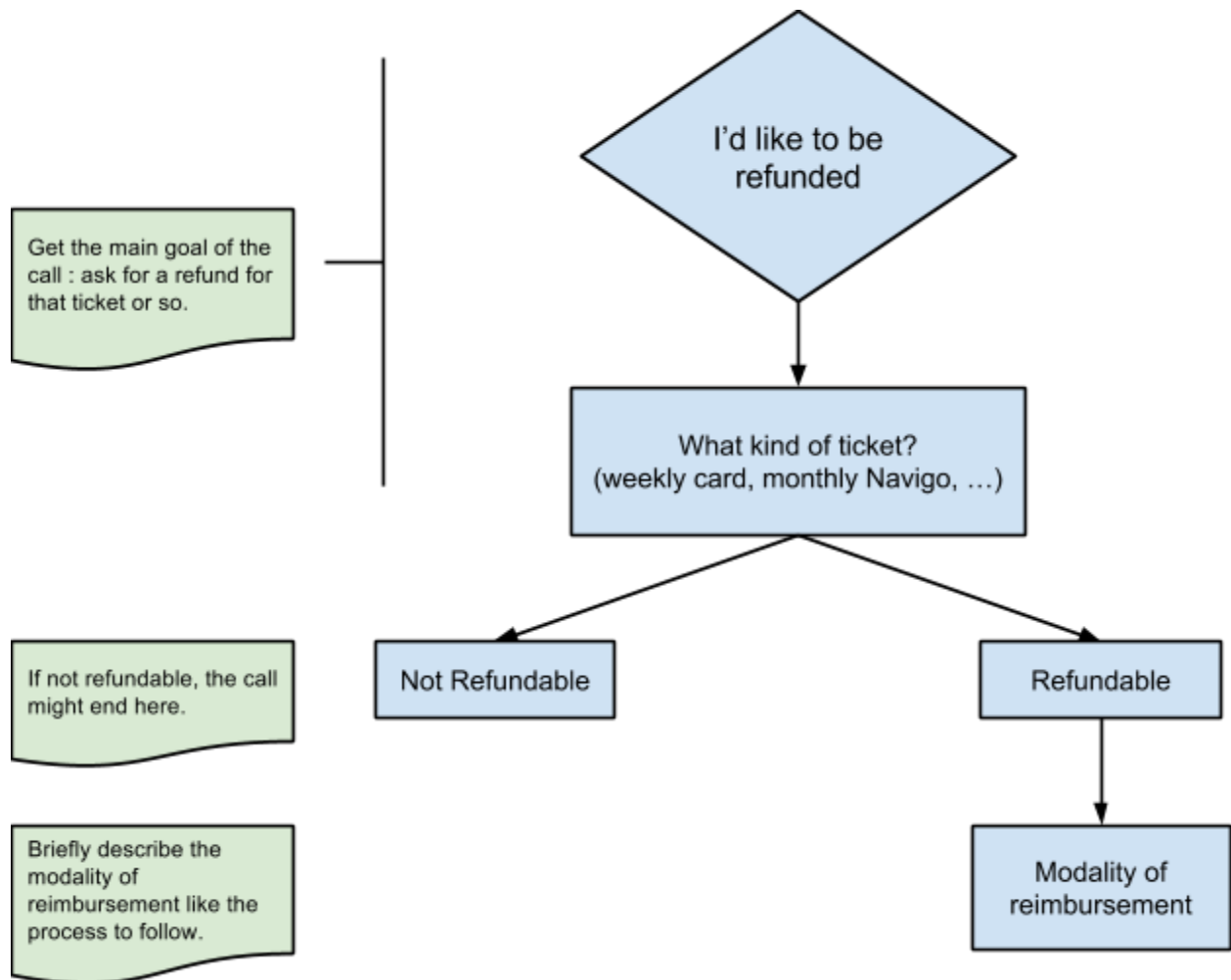
3 - Official report



Official report call are frequently pretty long (to explain everything or so).

Basically we try here to sum up the report in one or two sentences and then briefly give the solution given by the adviser (like "communication of the right service")

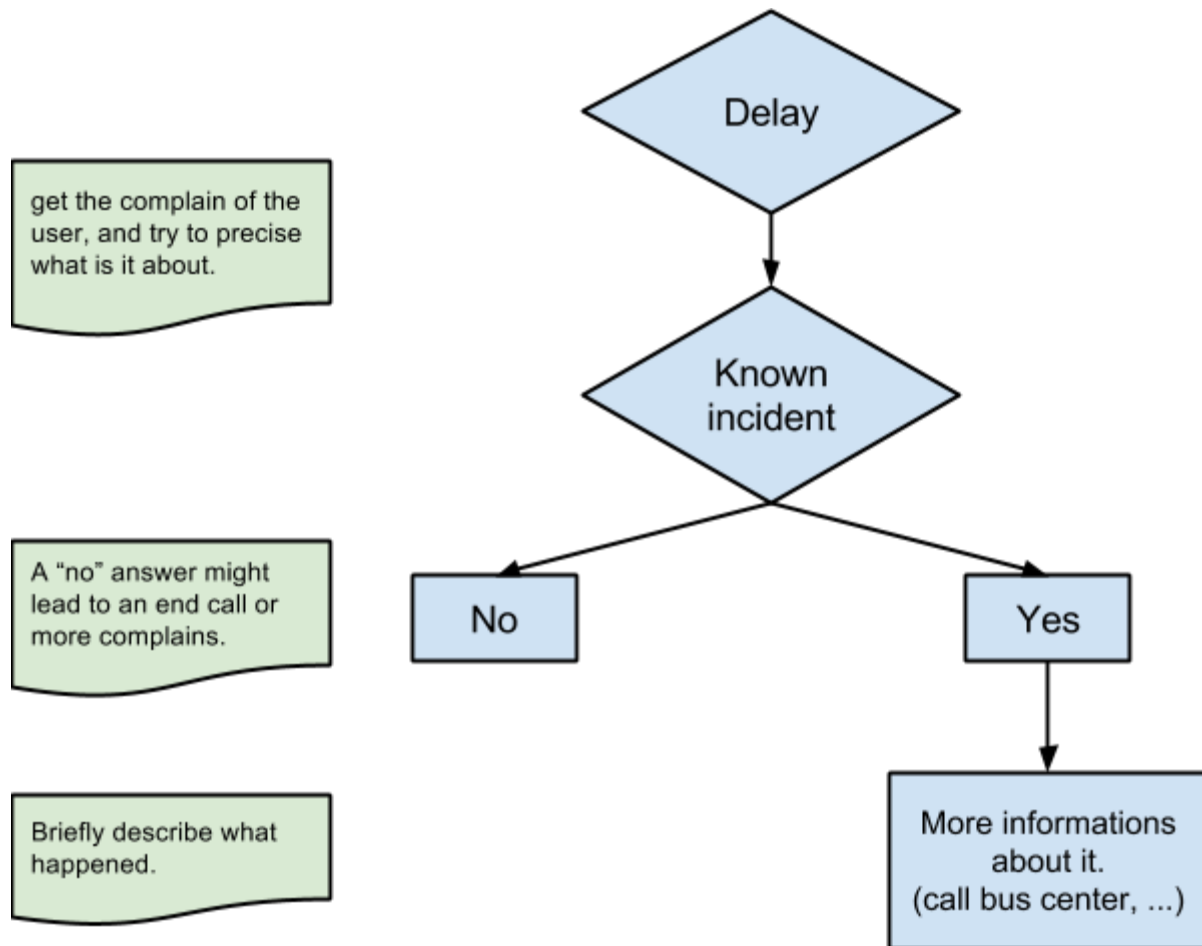
4 - Reimbursement



There is a lot of misunderstanding in these kind of call, depending on the knowledge of the caller, but because of the variety of the cards/contracts/errors/... the exchange between the caller and the adviser are pretty numerous.

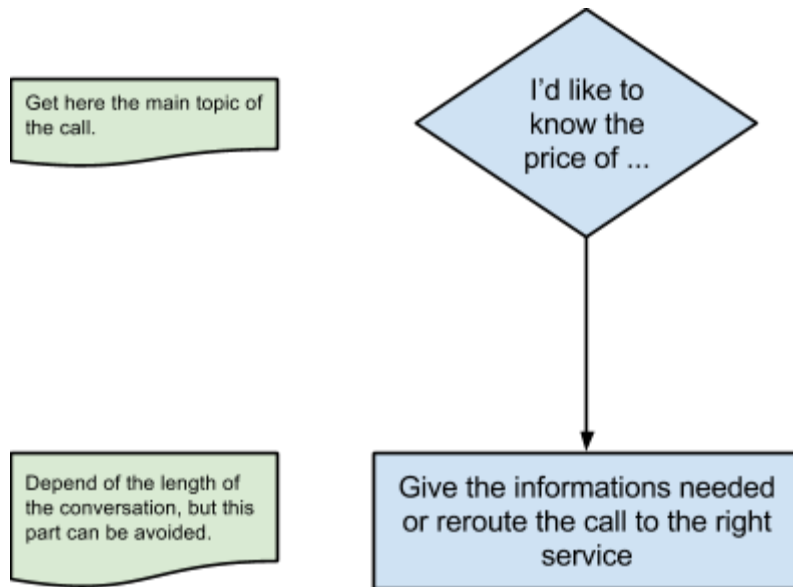
So just get the main reimbursement issue summed up and then as for the other scenario get the solution if it's relevant.

5 - Delay, incident



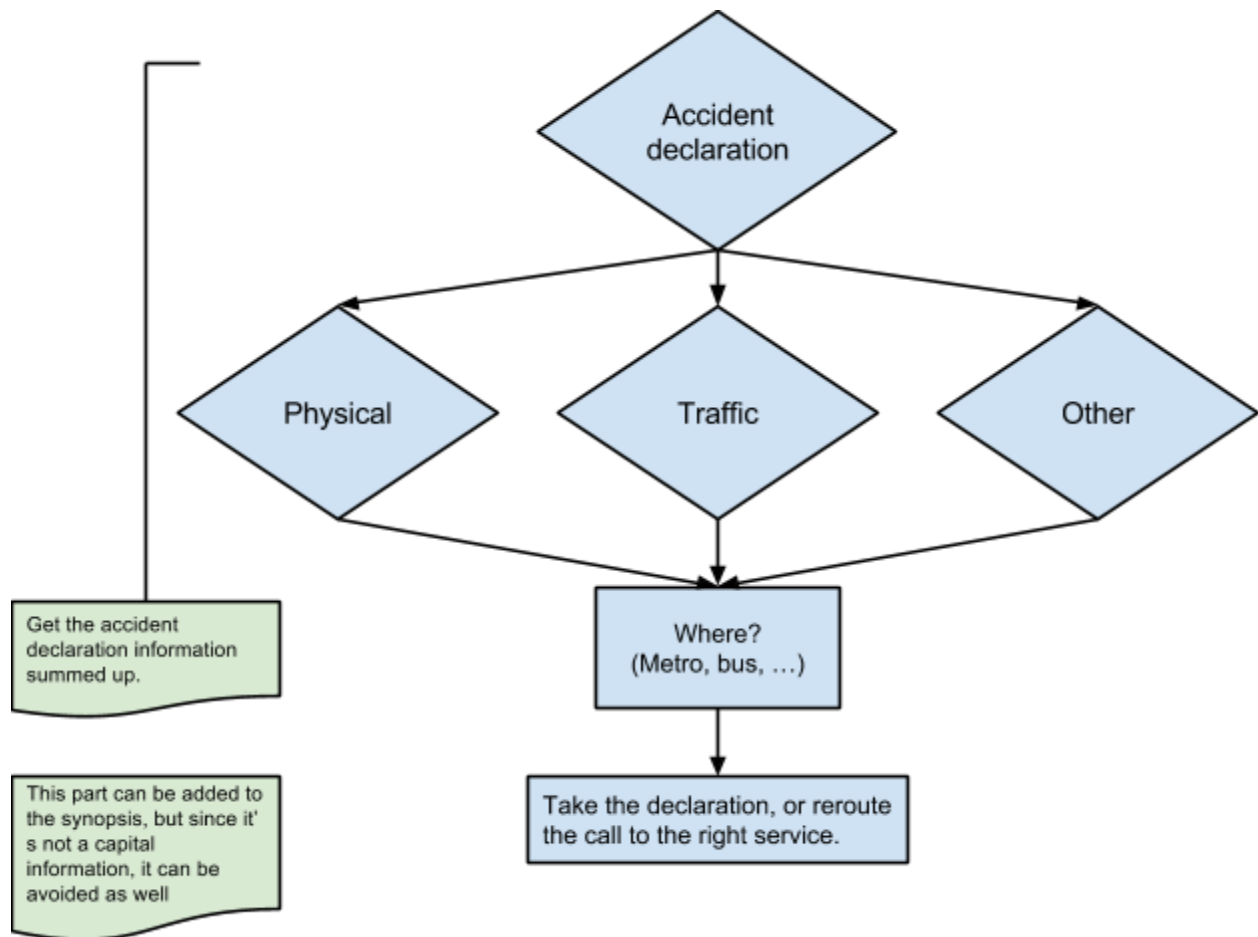
In these calls the adviser will frequently call the bus center or another service to ask for more information. Generally we don't really consider that call in call, but we prefer to sum up the whole thing just with the final answer/solution to give to the caller.

6 - Prices



Pretty much the same cases that in the previous scenario (5 - Delay, incident).
in some cases (the easiest and the shortest) we like to just sum up the whole conversation by the description of the issue (e.g. "asking for prices about an orange card")

7 - Accidents



Pretty much the same as 5 - *Delay, incident*. We sum up the declaration, and then get the solution/answer.