

ACCIDENT ROOT CAUSES IDENTIFICATION USING A TAXONOMY

JEROME MAGOUE¹, MATTHIEU PERIN² & SIMON COLLART-DUTILLEUL²

¹Vulcain Ingénierie, France

²Gustave Eiffel University, IFSTTAR/ESTAS, France

ABSTRACT

Railway accidents are rare, but for the sake of safety improvement they are all investigated by experts. Whereas this solution has proven its correctness to increase the overall safety in transportation, the lack of a methodological analysis formalization in case of similar accidents may lead to the non-detection of common root causes and to forfeit their removals. The aim of this work is to propose a formalized taxonomy based on the accident causes, classified by main categories such as “technical causes”, “human causes” or “organizational causes”. This classification is then used as a seed for safety analysis, like fault tree analysis, to detect and quantify possible common root causes for similar railway accidents. Along with this paper, the methodology is exemplified on a specific accident type: the derailments caused by over-speed in curves.

Keywords: taxonomy, railway, accident, root causes, fault tree analysis.

1 INTRODUCTION

Although railway accidents are subject to expert investigations, there is a lack of methodological analysis formalization in case of similar accidents, allowing an identification of the root causes by the mean of correlation. Actually, multiple occurrences of a given factor may lead to stronger conclusion and will allow the definition of corrective actions to avoid the considered class of accident.

The aim of this work is to propose a methodology to prepare taxonomy of the accidents causes, classified by main categories such as “technical causes”, “human causes” or “organizational causes”, and then use this result as a seed for safety analysis: root cause analysis and fault tree analysis (FTA).

First step of the method proposed is to build the taxonomy based on expert knowledge for main categories, and then refine in subcategories based on analysis of existing events. Using this taxonomy, a causal tree analysis is performed in order to track all the actions and conditions (or causal factors) that are necessary and sufficient for a given consequence to occur.

The second and last step is to setup safety analysis using our previous results. In this paper, a fault tree is built. Its main purpose is to help identifying potential causes of system failures. It should also be used to evaluate the probability of the top event using analytical or statistical methods. These calculations involve system quantitative reliability and maintainability information, such as failure probability, failure rate and repair rate.

Once completing the FTA, it is easier to focus the efforts on improving system safety and reliability on pinpointed core causal factors and their base event counterparts.

Along this paper, the methodology is exemplified on a specific accident: the derailments caused by over-speed in curve. For this case study, the available worldwide railway accident data are analysed in order to build taxonomy. The adequacy of the taxonomy is discussed with regards to the socio-technical context of accident reports. Indeed, fifty years old accident occurrences have to be analysed while keeping in mind the shift in scientific and technical knowledge. Cultural and legal evolution is something to be taken into account too when reviewing old accident reports.



As prospects, consulting the scientific state of the art in order to propose further formalization of the methodology is proposed.

2 BUILDING THE TAXONOMY

2.1 Approach for preparing the taxonomy

2.1.1 Definition of taxonomy

Taxonomy is a term issued from biology. It is the science of naming, describing and classifying organisms and includes all plants, animals and microorganisms of the world. Using morphological, behavioural, genetic and biochemical observations, taxonomists identify, describe and arrange species into classifications, including those that are new to science. Carl Linnaeus (1707–1778) is known as the father of taxonomy.

In the proposed methodology, this approach is transposed to perform the identification of root causes of train accidents.

As a minimum, train accident root causes can be classified into three main categories:

- Technical causes;
- Human causes;
- Organizational causes.

Each main category will be divided into sub-categories, as per the results of the analysis. Any new main category could be added if it is highlighted by the analysis.

2.1.2 Steps proposed to prepare the taxonomy

In order to prepare the taxonomy, the following main steps are performed:

- Definition of the set;
- Exhaustive analysis of the set;
- Definition and classification of the categories and subcategories.

Each activity will be detailed in the following sections.

2.1.3 Presentation of the example

We will illustrate how the methodology is applied by using a type of train accident: derailment in curve due to overspeed. This kind of accidents occurs more than 20 times in the last century.

2.2 Definition and analysis of the representative set of events

2.2.1 Definition of the set

A representative sample is a small quantity of something that accurately reflects the larger entity. An example is when a small number of people accurately reflects members of an entire population. In a classroom of 30 students, in which half the students are males and half are females, a representative sample might include six students: three males and three females.

When a sample is not representative, the result is known as a sampling error. Using the classroom example again, a sample that includes six students, all of whom are males, would not be a representative sample.

A representative sample parallels the key variables and characteristics under examination. Some examples include sex, age, education level, socio-economic status or marital status. Using a larger sample size increases the likelihood that the sample more accurately reflects



what actually exists in the population. Any information collection with biased tendencies is unable to generate a representative sample.

A representative sample allows the collected results to be generalized to a larger population. For most marketing or psychology studies, it is impractical in terms of time, finances and effort to collect data on every person in the target population. This is especially impractical for large population such as an entire country or race.

The use of sample groups poses risks, as the sample may not accurately reflect the views of the general population. One of the largest risks is developing a sample that is not truly representative. This most likely occurs because the population group is too small. For example, when comparing data relating to gender, a representative sample must include individuals of different ages, economic status and geographical locations. Such information typically requires a diversification of information-collecting sites.

Random sampling involves choosing respondents from the target population at random, to minimize bias in a representative sample. While this method is more expensive and requires more upfront information, the information yielded is typically of higher quality. Purposive sampling is more widely used and occurs when the managers target individuals matching certain criteria for information extraction. Ideal interview candidates receive profiles. Although this leads to the potential of bias in the representative sample, the information is easier to collect, and the sampler has more control when creating the representative sample.

2.2.2 Application

Table 1 provides the list of train derailments in curve. This list has been prepared using existing accidents database such as the European Railway Agency accidents base.

For each event, it is assessed if the available input information allows an exhaustive analysis of the accident in order to go deeply into the root causes.

2.3 Definition of categories

2.3.1 Listing categories and subcategories

For each item of the representative sample, a review is performed by an expert of the topics to be analyzed. The expert will identify the categories that meet the targets of the review. He is free to use any available input.

Categories and subcategories should not be defined in advance, they will be identified throughout the review process without any limit.

When each item of the representative sample has been reviewed and categories have been defined, the taxonomy can be defined.

2.3.2 Application

As explained in Section 2.2.2, the events to be analysed are chosen based on the available inputs. Main inputs used are the accident reports (as far as available, official reports following investigations).

These reports are analysed by a railway safety expert whose target is to identify the root causes of the accidents. It has to be reminded that investigation techniques are the ones used when the reports were issued.

The main target of this review is to focus on the analysis on the facts provided in the reports. Conclusions of the reports are only checked at the end in order to compare the results with the ones from the railway safety expert.



Table 1: Set of events.

Main inputs used	Date	No. of fatalities	Comments	Place of accidents
[1]	01/07/1906	28	–	Salisbury, UK
[2], [3]	01/11/1918	98	–	New York, Malbone Street, USA
[4], [5]	1947	24	Not analyzed because of the lack of input information	Pennsylvania, USA
	1947	16	–	Camp Mountain, Australia
–	02/10/1947	184	–	Hachiko Line, Japan
[6]	1955	17	–	Sutton Coldfield, UK
[7], [8]	1969, 1984, 1994	6	Six fatalities in three separate accidents	Morpeth, UK
[9]	1972	6	–	Eltham Well Hall, UK
[10], [11]	1989, 1996	8	Eight fatalities in two separate accidents	Cajon Pass, USA
–	1997	8	Not analyzed because of the lack of input information	Piacenza, Italy
[12]	2000	9	–	Bruhl, Germany
[13], [14]	31/01/2003	7	–	Waterfall, Australia
[15]	15/11/2004	0	–	Cairns Tilt train, Australia
[16], [17]	25/04/2005	107	–	Amagasaki, Japan
[18]	2006	41	–	Valencia, Spain
–	06/07/2013	47	Not analyzed because of the lack of input information	Megantic Lake, Canada
[19], [20]	24/07/2013	79	–	Santiago de Compostela, Spain
[21]	01/12/2013	4	–	Spuyten Duyvil, NY, USA
–	09/11/2014	0	Not analyzed because of the lack of input information	Colebrook, Tasmania
[22]	12/05/2015	8	–	Philadelphia, USA
[23]–[25]	14/11/2015	11	–	Eckwersheim, France

Table 2 summarizes the root causes identified by the railway safety expert for three representative accidents. These accidents have been chosen among the 20 others because of the involvement of the human and the organisation in the main causes of the accidents.

As a result of the expert review, root causes have been sorted into three main causes' categories: Infrastructure/Hardware, Human, Society. This classification is the taxonomy.

Tables 3 and 4 provide for the taxonomy for human and society route causes.



Table 2: Results of the analysis.

Location of accidents	Root causes identified
Amagasaki, Japan	<ul style="list-style-type: none"> • Horizontal alignment – strong radius of the curve • Driver error (overspeed, no respect of signaling information) • Lack of driver's knowledge or training • Lack of driver's experience • Lack of conscience of risks • Problem of organization • Pressure of organization
Santiago de Compostela, Spain	<ul style="list-style-type: none"> • Horizontal alignment – strong radius of the curve • Lack of safety signaling system • Driver error (overspeed, no respect of signaling information) • Error in procedures • Lack of conscience of risks • External perturbations
Eckwersheim, France	<ul style="list-style-type: none"> • Horizontal alignment – strong radius of the curve • End of line, connection between two lines • Driver error (overspeed, no respect of signaling information) • Error in procedures • Lack of driver's knowledge or training • Lack of driver's experience • Lack of conscience of risks

Table 3: Taxonomy – Human causes.

Factors	Code
<i>Driving error</i>	<i>H01</i>
Driver error	H01.01
Error in procedures	H01.02
<i>Knowledge/training</i>	<i>H02</i>
Lack of driver's knowledge or training	H02.01
Lack of driver's experience	H02.02
<i>Human behaviour</i>	<i>H03</i>
Use of alcohol, drugs	H03.01
Lack of conscience of risks	H03.02
Driver's health problem/personal problem	H03.03
External perturbations	H03.04

Table 4: Taxonomy – Society causes.

Factors	Code
<i>Organization issue</i>	<i>S01</i>
New organization	S01.01
Problem of organization	S01.02
Pressure of organization	S01.03
<i>Strike in the company</i>	<i>S02</i>
<i>Behavior of travellers</i>	<i>S03</i>
<i>Competition</i>	<i>S04</i>

2.4 Finalization of the taxonomy

The first revision of the taxonomy is prepared based on the expert review of the representative set of events.

It could be decided to use the taxonomy as it is or to try to develop the first results to make it as exhaustive as possible. The possible solutions to do so are:

- To use dependability tools (such as causal tree or FTA) to identify an exhaustive list of scenario;
- To review and complete the taxonomy during a workshop organized with several experts with experience on similar or connected subjects (will be detailed in a future publication) [27];
- To extend the representative set of events (will be detailed in a future publication).

Section 3 provides details on the way to use and enhance the taxonomy.

3 USING THE TAXONOMY RESULTS

3.1 Statistical approach

Once quantitative data have been collected, statistical analysis can be carried out in order to make sense, and draw some inferences from the data.

There is a wide range of possible techniques that can be used [26].

The first thing to do with any data is to summarise it, which means to present it in a way that best tells the story.

One of the most common techniques used for summarising is using graphs, particularly bar charts, which show every data point in order, or histograms, which are bar charts grouped into broader categories.

In our analysis of 20 accidents cases, first conclusions are:

- More than 75% of the accidents are caused by a human error;
- Among this 75%:
 - 80% are caused by lack of training, knowledge or experience;
 - 66% are caused by under estimation of the possible risk;
 - 33% are caused by errors in procedures;
- Only 25% of the accidents are caused by rolling stock failures;



- 55% of the accidents are caused by issue with the organization, and among this 55% because of pressure on the driver.

As a first conclusion of this statistical analysis, if we want to decrease the number accidents, we have to focus on the driver itself in order to provide the adequate level of training and safety systems onboard to assist in the driving activities, especially during stressful conditions.

3.2 Causal factor tree analysis

3.2.1 Listing the categories and subcategories

Causal factor tree analysis is a root cause analysis technique used to record and display, in a logical, tree-structured hierarchy, all the actions and conditions (or causal factors) that were necessary and sufficient for a given consequence to have occurred.

This technique:

- Provides structure for the recording of evidence and display of what is known.
- Through application of logic checks, gaps in knowledge are exposed.
- Tree structure is familiar and easy to follow.
- Can easily be extended to handle multiple (potential) scenarios.
- Can incorporate results from the use of other tools.
- Works well as a master investigation/analysis technique.

In the contrary, this

- Cannot easily handle or display time dependence.
- Sequence dependencies can be treated, but difficulty increases significantly with added complexity.
- Shows where unknowns exist, but provides no means of resolving them.
- Stopping points can be somewhat arbitrary.

3.2.2 Application

All the categories and subcategories identified in the taxonomy have been integrated in the causal tree (Fig. 1) in order to show all possible scenario.

This figure allows to check the overall consistency of the model and the possible lack in the taxonomy [28].

3.3 Fault tree analysis

3.3.1 Listing the categories and subcategories

FTA is a top down, deductive failure analysis in which an undesired state of a system is analysed using Boolean logic to combine a series of lower-level events. This analysis method is mainly used in the fields of safety engineering and reliability engineering to understand how systems can fail, to identify the best ways to reduce risk or to determine (or get a feeling for) event rates of a safety accident or a particular system level (functional) failure.

FTA methodology is described in several industry and government standards, including NRC NUREG-0492 for the nuclear power industry, an aerospace-oriented revision to NUREG-0492 for use by NASA, SAE ARP4761 for civil aerospace, MIL-HDBK-338 for military systems, IEC standard IEC 61025 is intended for cross-industry use and has been adopted as European Norm EN 61025.





Figure 1: Extract of the causal tree for “Derailment in curve due to overspeed”.

Any sufficiently complex system is subject to failure as a result of one or more subsystems failing. The likelihood of failure, however, can often be reduced through improved system design. FTA maps the relationship between faults, subsystems, and redundant safety design elements by creating a logic diagram of the overall system.

The undesired outcome is taken as the root (“top event”) of a tree of logic. Working backward from this top event we might determine there are two ways this could happen: during normal operation or during maintenance operation. This condition is a logical OR. Considering the branch of occurring during normal operation perhaps we determine there are two ways this could happen: the press cycles and harms the operator or the press cycles and harms another person. This is another logical OR. We can make a design improvement by requiring the operator to press two buttons to cycle the machine – this is a safety feature in the form of a logical AND. The button may have an intrinsic failure rate – this becomes a fault stimulus we can analyze. When fault trees are labeled with actual numbers for failure probabilities, computer programs can calculate failure probabilities from fault trees. When a specific event is found to have more than one effect event, i.e. it has impact on several subsystems, it is called a common cause or common mode. Graphically speaking, it means

this event will appear at several locations in the tree. Common causes introduce dependency relations between events. The probability computations of a tree which contains some common causes are much more complicated than regular trees where all events are considered as independent.

The tree is usually written out using conventional logic gate symbols. A Cut Set is a combination of events, typically component failures, causing the top event. If no event can be removed from a Cut Set without causing the top event, then the Cut Set is called a Minimal Cut Set.

3.3.2 Application

A fault tree has been prepared with the following top event “derailment in curve due to overspeed”. A sample of this tree is shown in Fig. 2.

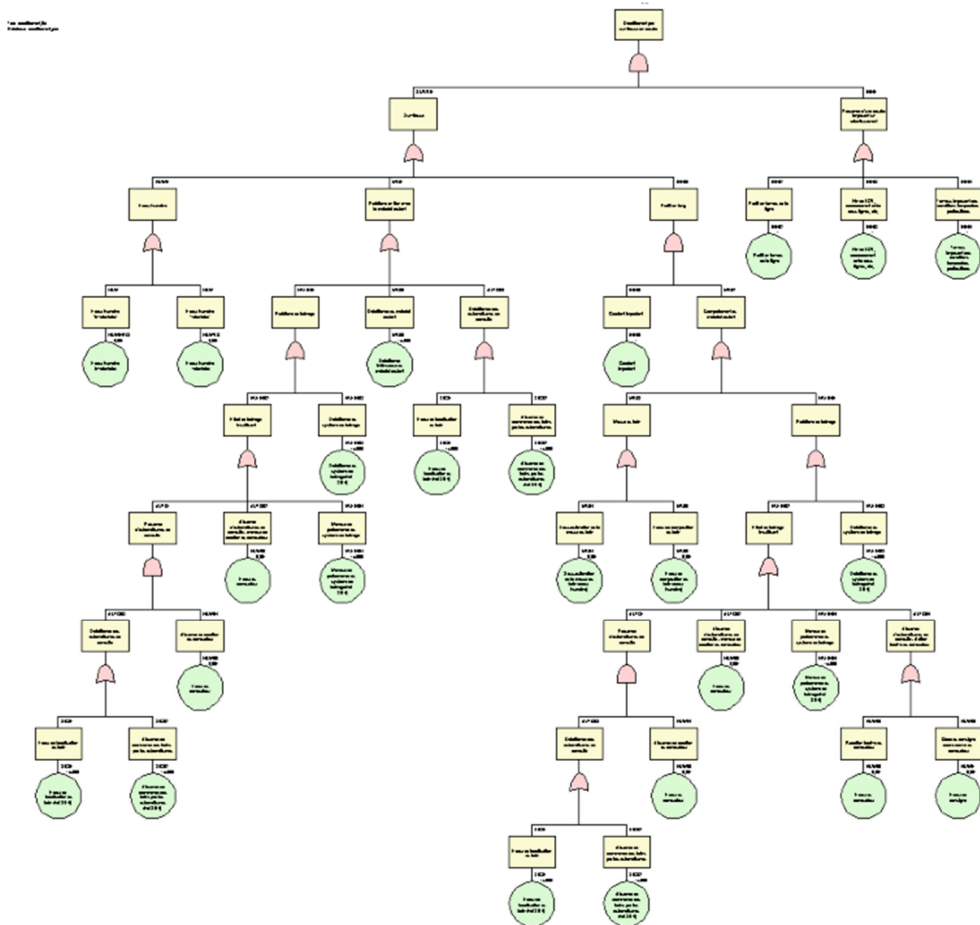


Figure 2: Fault tree for “Derailment in curve due to overspeed” (extract).

This tree could be used:

- To define the probability of the top event;
- To define the possible common cause failures;
- To define the minimal cut sets;
- To propose enhancement to reduce the overall probability of the top event.

4 CONCLUSIONS

This article provides the demonstration that it is possible to combine the taxonomy approach with traditional dependability tools to analyse and determine common root causes of a list of similar accidents.

The proposed short term development of this methodology could be:

- To split causes between direct and indirect causes... and to assess the ratio between each kind of causes;
- To use the analysis results to perform simulation of accidents;
- To apply the methodology on other type of accidents;
- To cross check the above results with the accidents taxonomy prepared by the European Railway Agency [29].

Considering the knowledge engineering perspectives, the taxonomy can be handled using an ontology-based framework like Protégé. The Ontorail (<https://ontorail.org>) initiative is clearly helping, but it only mainly defines infrastructures. Rolling Stocks automatism and operating rules are still not considered.

Using conceptual computing will probably need strong analysis efforts before providing efficient results. Anyway, using GORO, the basic principle that can be found in accident may be exported to quite different scenarios [30]. It is to be explained that this conceptual modelling is the correct answer to the following situation: accidents are rare, so the time range of the analysis is wide, whereas the technological context is rapidly evolving. “Technologies changes, but concepts remain”, conceptual modelling seems to be a good tool to apply this principle.

REFERENCES

- [1] Pringle, J.W., Board of Trade, Accident returns: Extract for accident on Salisbury on 1st July 1906, 1906.
- [2] www.nycsubway.org, Malbone Street wreck (New York Times, 1918).
[www.nycsubway.org/wiki/Malbone_Street_Wreck_\(New_York_Times,_1918\)](http://www.nycsubway.org/wiki/Malbone_Street_Wreck_(New_York_Times,_1918)).
- [3] Cudahy, B., *The Malbone Street Wreck*, Fordham University Press: New York, 1999.
- [4] Queensland Railway’s Interest Group, Camp Mountain disaster: Labour Day 1947. www.qrig.org/articles/camp-mountain-disaster-labour-day-1947.
- [5] Queensland Parliamentary Debates, Legislative assembly, Thursday 20 Nov. 1947, pp. 1528–1530, 1947.
- [6] UK Ministry of Transport, Report on the derailment which occurred on 23/01/1955 at Sutton Coldfield
- [7] UK Ministry of Transport, Report on the derailment that occurred on 7 May 1969 at Morpeth.
- [8] UK Ministry of Transport, Report on the derailment that occurred on 24 Jun. 1984 at Morpeth.
- [9] UK Ministry of Transport, Report on the derailment that occurred on 11 Jun. 1972 at Eltham (Well Hall) Station.



- [10] NSTB, Railroad accident report: Derailment of southern pacific transportation company freight train on main 12, 1989.
- [11] National Geographic, Crash scene investigations: Runaway train (San Bernardino disaster), (Video).
- [12] Eisenbahn-Bundesamt (EBA), Untersuchungsbericht: Entgleisung des D203 im Bahnhof Brühm am 06.02.2000.
- [13] Special Commission of Inquiry into the Waterfall Rail Accident, Final report, Volume 1, Jan. 2005.
- [14] Special Commission of Inquiry into the Waterfall Rail Accident, Final report, Volume 2, Jan. 2005.
- [15] Australian Transport Safety Bureau and Queensland Transport Rail Safety Investigation Report QT1472 into the derailment of Cairns tilt train VCQ5 north of Berajondo on 15 Nov. 2004.
- [16] Aircraft and Railway Accidents Investigation Commission, Railway accident investigation report: Train derailment accident between Tsukaguchi and Amagasaki stations of the Fukuchiyama Line of the West Japan Railway Company.
- [17] National Geographic, Seconds from disaster: Runaway train (Amagasaki disaster), (Video).
- [18] Comision de investigacion del accidente de la linea 1 de metrovalencia de 3 de Julio de 2006.
- [19] Informe final sobre el accidente grave ferroviario, 0054/2013 ocurrido el dia 24.07.2013 en las proximidades de la estacion de Santiago de Compostela.
- [20] Railvolution, Tragic derailment at Santiago de Compostela, Volume 13, No. 4.
- [21] National Transportation Safety Board, Railroad accident brief: Metro-North railroad derailment.
- [22] National Transportation Safety Board, Derailment of Amtrak passenger train 188, Philadelphia, Pennsylvania, 12 May 2015.
- [23] SNCF, Direction Générale de l’Audit et des Risques: Rapport d’enquête immédiate no. 2015-AS-67 – Déraillement du Train d’essai no. 814521 sur le raccordement de Vendenheim de la LGV Est Européenne en construction le 14 Nov. 2015.
- [24] BEATT, Note d’étape sur le déraillement de la rame d’essai de la LGV Est-Européenne survenu le 14 novembre 2015 à Eckwersheim.
- [25] BEATT, Rapport d’enquête technique sur le déraillement d’une rame TGV lors d’une marche d’essai sur la LGV Est-Européenne le 14 novembre 2015 à Eckwersheim (67), 2017.
- [26] Wilson, P.F., Dell, L.D., Anderson, G.F., *Root Cause Analysis: A Tool for Total Quality Management*, ASQ Quality Press: Milwaukee, WI, pp. 8–17, 1993.
- [27] International Electrotechnical Commission, Fault Tree Analysis, Edition 2.0, EC 61025, 2006.
- [28] Suryanto, H. & Compton, P., Learning classification taxonomies from a classification knowledge based system, University of Karlsruhe.
- [29] European Railway Agency, Designing the common occurrences and taxonomy for Common Occurrence Reporting Programme ERA-PRG-004-TD-002 V 2.0.
- [30] Negri, P.P., Souza, V.E.S., de Castro Leal, A.L., de Almeida Falbo, R. & Guizzardi, G., Towards an ontology of goal-oriented requirements. *CibSE*, pp. 469–482, 2017.

