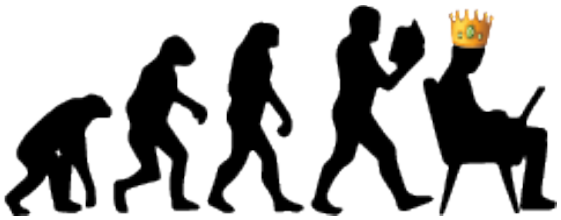
Language in cyberspace

By Inge Wertwijn, S1033944, 2967 words

Introduction

Cyberspace may turn out to be man's finest achievement. It has provided us with a world we can enter into at will – anytime and anyplace. In only 20 years' time, it has changed our lives. We use it to store, retrieve and distribute information – text, image, sound or video. We use it for interaction, both socially and for work. It is a marketplace where you can buy anything you want at knockdown prices. It provides a myriad of services: filling out a tax form, ordering a taxi, voting, designing a building, playing a game, doing a statistical analysis, learning a language, gambling or finding out tomorrow's weather forecast. We use it to instruct machines to perform all sorts of tasks for us: printing spare parts, opening and closing bridges, instruct our smart-home appliances such as fridges, lights, heating and even the vacuum cleaner. Perhaps the most important characteristic of cyberspace is that it spans the globe. Through it, we are able to reach any number of people or devices connected to it; simultaneously if we want. The future is even brighter: already drones and robots are fit for simple commercial use, decision making is more and more left to algorithms, and futurologists have predicted that by 2050 we will be able to upload our mind into cyberspace, and then download it again into a new body.



However, there is a downside to these wonderful new possibilities. Digital reality is much easier to manipulate than physical reality. With the right skills, it is a simple matter to steal information, change the way a program works, or tamper with a device. The ensuing possibilities for crime are endless. Steal an identity and use it to commit crimes. Create a botnet army out of virus-infected computers and use it to launch denial-of-service attacks on governmental services. Tamper with an ATM so that victims will hand over both their cash card and pin code. Create some fake news. Program a backdoor into a financial service and build up a nice little nest on an offshore account. Find con-victims on social media. Fake a little crypto coin mining. Sell copyrighted films, books and music. Spy on people using smart devices, then blackmail them. Profits are huge, 20 to 40% more than ordinary crime, even for a beginner. With care, it is simple to evade detection. Plus, costs are minimal. Looking at cyberspace from the point of view of a criminal, it is paradise. It is no surprise then, that cybercrime has generated much interest, not only from ordinary criminals but also from nation states, corporations, hacktivists, cyberterrorists and even script kiddies showing off their cleverness.

You might be thinking that surely the police and the government will protect your rights. You might believe that the organisations you hand your information to or whose services you use, will take care of your interest, if only because they don't want to lose clients. Possibly you even followed one of these Digi-aware courses and you know what is and what is not safe to do on the internet. Unfortunately, you are quite wrong. All this will only help you not to become an accidental victim. If you are targeted, of simply unlucky, you really don't stand a chance. You will be come a victim, in a big or small way. Do you know why? Because in technology, the human link is always the weakest. Cybercriminals will always have more money, more resources and more time on their hands than regulatory forces such as the police or the secret service.

Rules and regulations to the rescue

Let's take stock. So far, not so good. As a species, we have become totally dependent on cyberspace and we are entrusting it with almost everything we depend upon. That same cyberspace has become very attractive to criminals, who know their way around it even better than we do. So what is being done about this? There is only so much an individual can do to protect him or herself. We have to look elsewhere for help.

Countries, organisations, institutions: they all are expected to take care of their assets, including digital assets such as information and services. Protection is especially necessary when damage or misuse has negative consequences for the public or the state. Assets need protection against many situations, ranging from range from common theft to a disgruntled employee bent on revenge; from industrial espionage to natural disaster; from human error to terrorist attack. In general terms, protecting digital assets means ensuring their availability, integrity and confidentiality up to a pre-agreed level. On this subject, in the past 20 years a multitude of (inter)national regulations have emerged, and more appear every day. These regulations guide, direct or impel companies to institute digital asset protection and to report on the level of compliance achieved.

In 2018, every minute of every hour:

- businesses spent $171,233 to defend themselves;
- $1,138,888 was lost to cybercrime
- 1861 people fell victim.

In the next years, these figures are expected to double, triple, even quadruple.

Failing to comply may be punished in various ways: a formal warning, a fine, a revoked licence, or public shaming; and may result in the loss of a job, bankruptcy or even a prison sentence. Many governmental and commercial organisations actually want implement security, because it is in their own interest to do so. However, there is a problem. These regulatory texts are hard to understand, and their meaning is often open to different interpretations. In practice, different interpretations cause arguments amongst security practitioners, with management, with auditors and legal regulators. Inevitably, this leads to less rather than effective security.

What do security regulations look like?

The characteristics of these regulations provide us with some clues about the underlying causes of the interpretations problems. Regulations are always in written form, containing a mix of persuasive, informative, descriptive and instructive texts. The contents are focussed on a specific subtopic within the field of digital security, rather than about the whole field. They are intended to regulate behaviour, typically containing a lot of must's, should's and oughts's. Regulations are issued by a high-level body, such as a government, a board of directors of an (inter) national organisation. They are authoritative, either as an official directive or regarded as a de facto standard. There is always a formal creation, publishing and maintenance process through which the regulations are made available to a large audience, usually the public and may or may not require payment. The regulatory text itself is always produced as a group effort, usually involving stakeholders, experts and policy makers. Typically, there is no mention of the author(s). Examples of security regulations are: General Data Protection Act (GDPR)(published by the European Commission), the ISO/IEC 27K family of standards on information security, published by the ISO/IEC Joint Technical Committee, particularly the ISO27001 and the ISO27002; both European standards; and the Baseline Information security Overheid (BIO), published by the Dutch Government.

Organisations tend to treat these regulations as a single point of truth, taking texts as literally as possible. They do this, because they must demonstrate compliance with these regulations. For the same reason, implementation is usually achieved through a top-down chain of command.

The art of misunderstanding

As we saw, regulations are riddled with meaning problems You might think, why should that be a problem? General wisdom dictates that if you don't understand something, you should go and ask. Why does that not work here?

One reason is that there is no one to ask. There is no author to ask for clarification, nor is there an easily accessible expert group. An additional problem is that reaching out to the publisher of the regulation in question, must be done through proper channels, i.e. not something just any employee of any organisation can do. Usually, the best that may be achieved is to send in a formal request for clarification - which may or may not be processed during a future maintenance window. Another reason is that readers tend not te be aware of the different meanings of a particular bit of text, because they assume that there is only one meaning, namely the meaning they have assigned themselves. Only when one reader happens to be confronted with a different interpretation by someone else, the initial assumptions may be questioned. Yet another reason is that no one likes to admit to a lack of understanding or knowledge. It is associated with losing face, particularly when the regulation in question is implemented from the top-down. Power and knowledge of important matters supposedly lives at the top, rather than in the workplace.

The nett result is that texts get interpreted in different ways by different people who all claim they are right even when they are working at cross purposes. This generally results in a confused implementation of the regulation, and ultimately, in compliance failure.

There are many causes which contribute to interpretation problems in these texts. However, let us begin with what, contrary to popular opinion, is not a cause. It is not the case that the authors of these texts are unable or unwilling to use plain language. Rather, they arrive at the final wording through a group effort. To achieve consensus, the outcome of a negotiation process, is much more important than clarity. Meaning problems which arise from this cause take the form of obfuscation and generally over-complicated text containing (too) many qualifiers and sub-clauses.

The same effect may be produced deliberately. Organisations that issue regulations are usually funded by public money and derive their status at least in part from their authority of being accepted by all parties involved. To keep that status and funding, they try to avoid any big confrontation with the intended audience. For that reason, expectations on compliance tend to be worded softly, so they won't chafe too much, allowing for an escape. This may be done by artfully introducing intentional vagueness into the text, for instance, by not being specific on whether something must, should or could be done.

Context is another issue. The same words will mean different things in different contexts, or to different people, and these meanings may even be contradictory. For instance, the term *special data* might be taken to mean data that need special care, or to data that are for some reason special. Yet the term also refers to data

which it is the special duty of the government to secure. Within the context of the General Data Protection Act it means something completely different again, namely data describing very particular human characteristics such as DNA, creed, race or political inclination.

Another example is the use of the word v*alue.* In Dutch governmental regulations the term refers to anything which, when compromised, will negatively affect the Dutch state or its partners. To security professionals, the term signifies the value of a company asset, expressed in either quantitative (money) or qualitative terms. In a business context, the term usually refers to the interest of an important stakeholder. In everyday speech, the term just means that the issue of some importance.

There also are knowledge problems which take various forms. There may be a lack of knowledge at the level of the intended audience. The committee or group composing the regulation may also have knowledge gaps. A knowledge gap may have an underlying cause, such as a belief about the extent to which it is possible or desirable to regulate behaviour, or an opinion about whether security threats are real or may be countered.

The language that is used to discuss digital assets and their security is another concern. Specialists need to express themselves in a way which another specialist on the other side of the world, in a different industry or with a different cultural background may understand. Within the field of information processing various modelling languages have been developed, ranging from formal, mathematical models to more descriptive languages. The creation of such languages are a rough implementation of a philosophical notion which was once very popular among English philosophers who believed in an ideal language. Descriptive languages have the added advantage of being designed to produce strong visualisations which can be shared with a less specialised audience. However, the problem with current 'descriptive' languages, is that the concepts they are built on, have been arrived at through trial-and-error and common sense. Inevitably concepts overlap, leave gaps, are overloaded or simply are not sufficiently clear for the use of capturing knowledge. In other words, the language is there, but its grammatical basis is unclear.

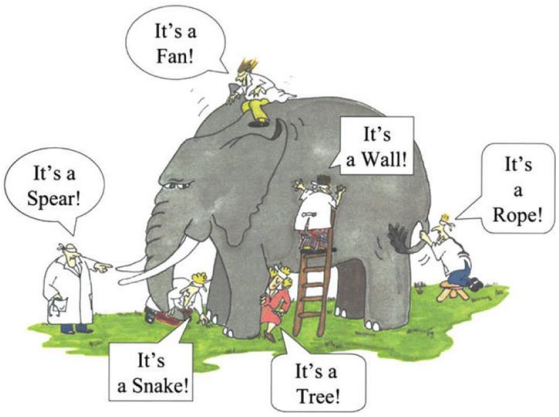
Much interest has centred on the possibility of capturing security concepts in formal taxonomies that semantic computer programs can process. This is another implementation of the "ideal language" idea, but more focussed on individual meanings rather than on the language structure. In principle, such a taxonomy works for all kinds of information, including security, and may be used to construct theories, harmonise concepts or create computer-based applications. Indeed, some real progress has been made in highly specialised sub-topics such as automatic threat detection in cyberspace. Yet that progress seems to have been possible only because there exists a straightforward cause-and-effect relation between a specific cyber-threat and the way to respond to it. This is not so simple for the rest of the security domain. Overall, security taxonomies for sub-topics are developed independently from each other. In a recent survey eight different families of security taxonomies were identified. Despite considerable work, these efforts do not converge but diverge. The lack of a common body of knowledge is seen as both a cause and a solution to the problem. Unfortunately, this means that in spite of all efforts, understanding of the interrelation of security concepts has not been increased in any usable manner.

The way forward

The above presents a general overview of problems encountered when interpreting regulations on cybersecurity and points to some possible causes. These causes may exist simultaneously and may interact. Much more work needs to be done on this to achieve a true identification of relevant causes and underlying factors. And there is more. Outside of the field of security, more language problems appear in our interaction with cybertechnology. There is

a general issue with the meaning of sentences that are exchanged, not in an individual interaction, but on social media, by proxy, through robots or by algorithms which produce human-like interactions. Currently there seems to be no theory of language that accounts for how we handle this. Yet in spite of us not knowing exactly how language works, how meaning is endowed, what it means to speak to each other and infer meaning, robots are currently being instructed to understand not just normal speech, but also indirect speech, such as "Where is the coffee" rather than "I want some coffee". Because that is the way we speak, and robots need to understand us. Already elderly people are being given such robots to keep them company. Sex robots that talk and interact are big business. In games, we talk to preprogrammed action figures. We may be chatting to a helpdesk, unaware that there is a chatbot, rather than a person, on the other side. All these are all examples of interaction with preprogrammed agents, where the intended meaning of what is said, is not produced by the "person" in front of you, but was put in long before the sentence ever reaches you. Such problems, where the original intention behind the text has become unavailable to us, are not dissimilar from the problems encountered when interpreting regulatory texts produced by anonymous bodies.

Language is all we have to instruct and guide each other. If it fails us when we try to protect our world and ourselves, the underlying problems must be sorted out. But by whom? Not by the people already entrenched in the everyday business of digital security. The situation is reminiscent of the fable of the six wise blind men and the elephant. The first man touches the elephant's trunk and says: "this is a thick snake". The second one reaches up to the elephant's ear and concludes: "this must be some kind of fan". The third pats its hide and claim: "a wall covered in leather, I am sure". The fourth winds the tail around its arm and says: "this is a fine rope". The fifth feels the side of the tusk and observes: "this is smooth and hard and pointy, it must be a spear". The final blind man embraces one of its legs and exclaims: "it's a tree!" None of them will ever conclude that this is an elephant.



Time for some good old fashioned thinking. Let's get the philosophers in to help. Point out the elephant in the room and say to them: it is high time that you stopped nit-picking over theories that applied to a world that is rapidly taken over by a new one. This is a paradigm shift. Give us a hand. Tell us in what situations it is a good idea to create an ideal language so that we may understand each other, and when it will only muddle things up. Develop a knowledge framework with a workable definition of trust that we may apply to digital agents, devices and technically enhanced humans. Help us to understand how to use words like "must", "should" and "ought" so that we may understand what responsibility they entail and how to apply these concepts to cyberspace. Work out the mechanism that people use for establishing common ground, so that we may use it make ourselves understood in this new technological era. Make us aware of the limitations of language interaction when it is not conducted face to face, so that so that we can learn to express ourselves without continuously misunderstanding each other. In short, create us a swiss army knife so we may know how use language to help us protect ourselves in these exciting but dangerous times.