

DEFINITIONS IN ONTOLOGIES

Selja Seppälä, Alan Ruttenberg, Yonatan Schreiber, Barry Smith

1. Introduction

Definitions vary according to context of use and target audience. They are to be made relevant for each context to fulfill their cognitive and linguistic goals. This involves adapting their logical structure, type of content, and form to each context of use. We examine from these perspectives the case of definitions in ontologies.

An ontology is a formal, machine-tractable representation of the types of entities and relations found in a given domain. Ontologies are built out of ontology elements called ‘terms’ (sometimes also called ‘classes’¹) and ‘relations’. Those elements are standardly associated with axioms and with linguistic and other information such as synonyms and unique alphanumeric identifiers for the terms, and associated notes or comments.² Importantly, ontologies — ideally, at least — contain definitions of their terms. What results when these elements are combined is a specification of a domain-specific vocabulary, comparable to a terminological dictionary but with additional formal structure able to serve computational reasoning.³

Ontologies have many applications,⁴ including: tagging large and heterogeneous bodies of data as a means of making the data available for integration, search, query, and analysis; use in natural language processing for semantically enriching texts; use in automated reasoning over annotated texts; use in decision-support systems; indexing of images for visually enriching ontologies; and linking of data of all types with relevant literature.

However, the more such applications are successful, the more we face the risks of knowledge silo formation driven by the accumulation of uncoordinated or inconsistent ontologies, undermining their success. To counteract these risks, ontologies must as far as possible be developed in a rigorous fashion in order that they will acquire the degree of trust from potential users that is needed to ensure that they will be used and re-used to the maximal possible degree.

¹ In some older traditions also called ‘concepts’ (Schulz, *et al.* 2013).

² See *RDF Schema 1.1*, <http://www.w3.org/TR/rdf-schema/>. See also Ceusters and Smith (2010), Rubin, *et al.* (2007), and Smith, *et al.* (2006).

³ For examples, see the biomedical ontologies at Ontobee (<http://www.ontobee.org>).

⁴ See for example, Bodenreider and Stevens (2006), Bucella, *et al.* (2011), Fonseca, *et al.* (2002), Hoehndorf, *et al.* (2015), Kontopoulos, *et al.* (2016).

One central strategy to achieve this effect is to ensure that ontologies incorporate both logical and natural language definitions of their terms. Unfortunately, the community of practice around ontology has, as yet, failed to develop a generally accepted theory of how definitions are to be authored and used in ontologies. Ontologists have also failed to develop a theory of the functions that definitions play. The result is that definition-authoring practices vary widely from one community to the next.

While textual definitions differ from logical definitions in a number of respects; they also share a number of characteristics. Our goal is to contribute to furthering the understanding of definitions in ontologies by (1) examining what characterizes the types of things called ‘definition’, (2) explaining the cognitive and linguistic functions of definitions, and (3) providing insight into the similarities and differences between textual and logical definitions in ontologies. Our analysis draws on the more detailed analysis of terminological definitions provided in Seppälä (2012, 2015) and takes into account the specific contexts in which ontologies are used.

2. A characterization of definitions

The term ‘definition’ can be used in a number of different ways, according to the different types of entities to which it refers. These include:

- a cognitive activity (the process of forming a certain cognitive representation);
- a cognitive representation (something in a person’s mind, analogous to a belief);
- a representational artifact (something shared that is deliberately created to do a certain job);
- an act of communication.

According to Robinson (1950: 13), the original use of the term ‘definition’ is to refer to a certain cognitive activity and application of the term to the other types of entities follows from this. The cognitive activity of defining consists in forming a mental representation that can serve as defining content about the definition’s object, usually preliminary to communicating this defining content to someone else. Such defining activity generally leads to a communicative act whose form and modality of expression will vary according to the communication situation.

The four entities in our characterization may be defined as follows:⁵

DEF1: *definition as a cognitive activity*

A cognitive activity performed by a cognitive subject that consists in forming a *cognitive representation* of some entity X and serves to specify what it is that makes an entity X rather than something else.

DEF2: *definition as a cognitive representation* (Smith, *et al.* 2006: 59, Smith and Ceusters 2015)

⁵ We extend the work presented in Smith, *et al.* (2006).

A cognitive representation composed of a set of items of knowledge or beliefs about a definition's object resulting from a *defining activity* in the sense of DEF1.

DEF3: *definition as a representational artifact* (Smith, *et al.* 2006: 59)

A representational artifact that expresses the *defining content* (DEF2) resulting from the *defining activity* (DEF1) and is communicated in a *defining act* (DEF4).

DEF4: *definition as an act of communication*

A communication act that consists of communicating *definition content* (DEF2) by means of a *representational artifact* (DEF3) to a receiver.

As it appears from these definitions, not only are the corresponding entities closely related to one another, they also presuppose the existence of other entities: participants (cognitive subjects), a communication situation (*in presentia* or *in absentia*), a modality of expression, and an object of definition (represented by the variable 'X'). In the following, we examine each definition in more detail.

2.1. DEF1: *defining as a cognitive activity*

In the sense of DEF1, defining is a cognitive process. It is an activity, performed by a definer, that is part of the more general activity of creation of knowledge-rich resources. Resource creation includes gathering specialized texts and information from experts to compile domain-specific corpora. Such corpora are then used for identifying definitional contexts — portions of the corpora that articulate items of knowledge or belief. Based on these contexts, the cognitive activity of defining starts.

From the identified definitional contexts, the definer selects relevant definitional content according to the target audience and the intended use of the definition — in this case, for construction of an ontology and its subsequent use by ontologists and domain experts. The intended use of the resource determines the extent to which the definition is to be descriptive or stipulative, as well as the logical form of the definition, as we will see in Section 3.2. With the selected definitional content, the definer constructs a definition (DEF2).

The resulting definition content is checked, for example, to ensure that there are no counter-examples and that it deals properly with problematic cases. Finally, the definer chooses one or more types of representational artifacts — linguistic, symbolic, depictive, etc. — that are adequate for rendering the definition content (DEF2) in the intended context of use. The resulting output is a definition in the sense of DEF3.

The defining activity presupposes the existence of one or more participants with specific roles. It involves at least one direct participant, a *definer*, who constructs a definition (DEF2) and a (real or imagined) *target audience* the definer has in mind. Both the definer and the target audience are persons who, as members of a particular community, have a specific role in that community. The definer can be a scientist, ontologist, terminologist, teacher, legislator, and so on; the target audience can, for example, consist of data curators, translators, learners, peers, and, as we will see,

computer systems. A third group of participants are the speakers (or users of language more broadly conceived), whose lexical use is captured in and communicated through a definition (Robinson 1950: 35-36). However, it seems odd to consider the latter group of individuals as participants in the defining activity; they are rather the source of the meanings expressed by the definitions.

In general, the agent of the defining activity can be considered either as an *immediate definer* or as a *mediator* (Chaurand and Mazière 1990: 272). The former corresponds to the agent who defines ‘freely’ or ‘naturally’ (Martin 1990, Rebeyrolle 2000) on the basis of her own knowledge and beliefs. The mediated definer is an agent who performs the defining activity either on the basis of direct consultation with or by using sources produced by the immediate definer. The immediate definer lays the basis from which the mediated definer can engage in the secondary activity of defining.

In the case of ontologies, the definer is generally an ontologist or a domain expert working in collaboration with an ontologist. Ontologists often perform the defining activity as mediators, applying a more or less established methodology that requires using primary sources, sometimes including other ontologies.

One thing that distinguishes ontologies from terminological dictionaries is that they are used not only by human beings but also by machines. Ontologies include textual as well as logical definitions, where, as in the words of Stevens, *et al.* (2011: 2), “textual definitions are human-facing, logical definitions are primarily machine-facing.” For example, *microbicide* in the Infectious Disease Ontology (IDO) is defined as follows:

A material entity with an antimicrobial disposition that is realized in a process of killing microorganisms.

```
SubClassOf
'material entity' and
(has_disposition some
  ('antimicrobial disposition' and
    (realized_by only (results_in some death))))
```

This has consequences for both the defining activity and its product. In the case of textual definitions, the contents and linguistic expression can be adapted to the target audience of the ontology by taking into account distinct receiver-profile types (for example, providing the relevant information for a given level of expertise and adapting the defining vocabulary accordingly). In the case of logical definitions, the cognitive representation (DEF2) has to be rendered to be read and interpreted by a machine to produce logical inferences.

2.2. DEF2: definition as a cognitive representation

The cognitive representation that is the output of the defining activity (DEF1) is also called a definition (DEF2). In this sense, ‘definition’ refers to the content that can be copied and concretized in various ways (see DEF3).

The cognitive representation is composed of *features*, i.e., pieces of knowledge/beliefs. Features represent properties of the object of the definition, that is, of things of a certain type or of a particular thing in the world, which is the focus of the activities and practices of domain experts and to which they refer with one or more specific term(s).

In domain-specific ontologies and terminologies, class terms denote either (i) a type of entities in the world (the instances), also called a ‘universal’, or (ii) a ‘defined class’, which denotes a group of entities demarcated by human beings on the basis of more or less arbitrary selection criteria (Arp, *et al.* 2015: 19). Examples of such defined classes are: American middle-class household, person who has been baptized a Lutheran, building located in the city of Buffalo, NY. The definition associated with either term describes features of the instances of the type or of members of the class.

Associated with a term is an ‘intended meaning’ (i.e., its semantic value) that is specified by its definition. This intended meaning consists of a set of two or more features (expressing items of knowledge or beliefs) that together represent some ‘portion of reality’ to which the defined term refers or is meant to refer (Ceusters and Smith 2015, Ceusters and Smith 2010, Smith, *et al.* 2006). These features form the contents of a definition in sense DEF2. For some communities, they may amount to nothing more than a description of what the defined term refers to; but to work in scientific and computational contexts, the relevant contents must amount to a statement of necessary and, whenever possible, sufficient conditions (see Section 3.4.).

Typically, a definition has the canonical form ‘X is a Y that Zs’, with the three-part structure:

1. a *definiendum* [‘X’], i.e., the defined term;
2. a *definiens* [‘a Y that Zs’], i.e., the part that expresses the definition content and that is called a ‘definition’ in ontologies and dictionaries;
3. a *copula* [‘is a’] that expresses an equivalence between *definiendum* and *definiens*.

In ontologies and dictionaries, *definiendum* and *definiens* appear in distinct ontology annotation properties and dictionary entry fields; the copula is usually implicit.⁶

In its concretized form (DEF3), the *definiens* consists, ideally, of a short sentence fragment provided as the object of a ‘definition’ field in a dictionary entry or annotation property value in an ontology. This kind of natural language definition is also found in specialized terminological dictionaries.

A good definition delimits the intended meaning of an ontology term by describing the instances of the type to which the term refers. It states that the Xs are of the type Y

⁶ Some ontologists use full form definitions that include the *definiendum* and the copula. Here, we are only concerned with the *definiens*.

and are distinguished from other instances of this type by some collection Z of one or more characteristic marks. For example, the Cell Type Ontology (CL) contains the following definition for the term *leukocyte*:

- (1) *An achromatic cell of the myeloid or lymphoid lineages capable of ameboid movement, found in blood or other tissue.*

This example shows that the term ‘leukocyte’ [X] refers to those things that are of the type achromatic cell [Y] and that are distinguished from other achromatic cells in virtue of being: of the myeloid or lymphoid lineages [Z₁]; capable of ameboid movement [Z₂]; and found in blood or other tissue [Z₃].

Here, the X part is classically called the species; the Y part (or ‘head’ of the definition), the genus and genus proximus when it is the immediate superordinate type; and the Z parts, the differentiae.

In the classical Aristotelian form, an ‘is_a’ or subtype relation between the species and genus is asserted or implied, as in example (1) above, which we read as: *a leukocyte is_a achromatic cell*.

A differentia may express any kind of relation relevant for describing and distinguishing the kinds of things to which the defined term refers (Smith, *et al.* 2005). In example (1) above, the relations expressed in the definition of *leukocyte* are respectively ‘develops_from’ (the myeloid or lymphoid lineages), ‘capable_of’ (ameboid movement), and ‘located_in’ (blood or other tissue). Together, the genus and differentia(e) parts of a definition thus constitute its internal semantic structure.

The logical form of a definition derives from the relationship between its intension (that which is said about the referent, i.e., a description of properties of the instances of the defined type) and its extension (the set of instances that fall under the intension). When definitions are viewed in these terms, we can distinguish four main logical forms:

– Classical definition: A definition where the intension holds for all instances of the type X that is defined and does not hold for any instance that is not of that type. In this case, the characteristics expressed by Y and Z are both individually necessary and jointly sufficient for something’s being an X. This type of definition, which forms the ideal recommended case, is also called a definition by necessary and sufficient conditions. A standard example of classical definition is that of a *triangle* as: *A rectilinear figure that has three sides*. (All triangles and only triangles are rectilinear and have three sides. Every figure that is rectilinear and has three sides is a triangle.)

A classical definition of the species-genus form (‘A is a B that Cs’) is also called an ‘Aristotelian definition’, illustrated already in *A man is a rational animal*. Here *man* is the species, *animal* is the genus (or parent type) and *rational* is the specific difference — it is that feature of an instance of the genus which makes it also an instance of the species. Here too, the definition may have multiple differentiae.

- Partial definition:⁷ A definition where the intension holds for all instances of the type that is defined but also holds for instances that are not of that type. A partial definition is a statement of necessary conditions that are not jointly sufficient. An example of partial definition is that of a *bird* as: *An animal that lays eggs*. (All birds lay eggs, but other animals also lay eggs.)
- Typical or prototypical definition: A definition where the intension holds for most of the instances of the type that is defined, especially the typical ones, but also for instances that are not of that type. An example of a prototypical definition is that of a *swan* as: *An aquatic bird with a long neck, usually having white plumage*, which holds for most swans and also for most *snow geese*.
- Instance definition or definite description (Russell 1905): A definition where the intension holds for only a single instance. This kind of definition would apply, for example, to resources that include what may be considered as proper names, such as the *Large Hadron Collider* (LHC) in a dictionary or ontology of nuclear physics. In this case, the relevant kind of differentiae would probably inform us about the geographical location of the LHC and specify that it is (or was until some point in time) “the world’s largest and most powerful particle accelerator.”⁸

Ideally, ontologies contain only classical definitions because their linguistic function is to disambiguate terms (see Section 4.2.). This is not to say that the other logical forms cannot appear, for instance in their textual definitions, but this is not ideal with respect to the function they are meant to fulfill in this context; without necessary and sufficient conditions, it becomes possible to interpret terms in a manner that does not conform to their intended use.

2.3. DEF3: definition as a representational artifact

The definition content (DEF2) is ultimately concretized or rendered as one or more representational artifacts (DEF3), which means it is given a tangible form that allows it to be communicated. The definition contents and its concretizations are then two sides of the same coin: they share the same structure and logical form — they can thus be analyzed in the same way, as we will see in Section 5.

2.3.1. Concretization forms

The concretization form of the definition content depends on the context of use and the relevant (material) support on which the definition is to be communicated in that context — paper, electronic. The concretized representational artifact can be linguistic or nonlinguistic.

Generally, when we use the term ‘definition’ in the sense of a representational artifact, we refer to a linguistic expression in some natural language. In this case, the

⁷ See, for example, <http://logic.stanford.edu/kif/definitions.html>.

⁸ Source: CERN, <http://home.cern/topics/large-hadron-collider>.

genus is generally expressed by means of a noun or verb, and the differentiae often by means of adjectives, relative clauses, and past or present participles. The form may vary according to specific languages, definition writing conventions, the editorial line of the resource, and target audiences.⁹ It may also be automatically checked (Seppälä 2006).

The defining content can also be concretized by other means, for example, in a graphical representation of the corresponding ontology branch formed by labeled arcs and nodes representing classes and the relations between them. Other examples of concretizations that go beyond natural language are logical formalisms, chemical symbols, and graphs.

In ontologies, definition contents are concretized in both linguistic and nonlinguistic forms for display on a computer screen and use by an automated reasoner: textual definitions are intended to be interpretable by human users, while logical definitions in the form of axioms are created, in part, so as to be able to be read and reasoned with by machines.

The axioms and graph in Figure 1 concretize the same definition content as the textual definition of *leukocyte* in example (1) above: *An achromatic cell of the myeloid or lymphoid lineages capable of ameboid movement, found in blood or other tissue.* These are all distinct forms of concretizations that allow for the communication of the same contents of a definition.

```

leukocyte
SubClassOf
  achromatic cell
  develops_from some hematopoietic stem cell
  capable_of some ameboidal-type cell migration
  part_of some immune system
  
```

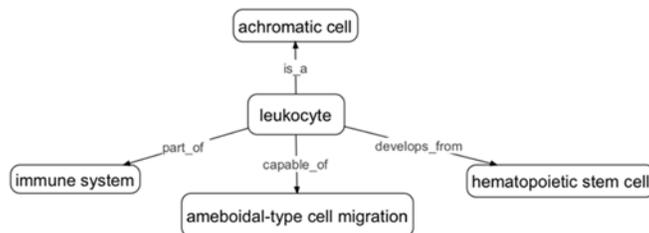


Figure 1: Alternative concretizations (axioms and graph) of the definition (DEF2) of *leukocyte*

⁹ Terminological manuals and guidelines state a number of general principles and recommendations relating to definition writing (ISO 704 2009, Pavel 2012, Pavel and Nolet 2001, Vézina, *et al.* 2009). For ontology-specific recommendations, see Arp, *et al.* (2015), Schober, *et al.* (2009), Seppälä, *et al.* (2014), Smith (2013).

2.3.2. Surface forms of a concretization type

Each type of concretization takes a surface form that depends on the target audience (including machines) and context of use.

The specific surface form of the linguistic (textual) concretization of the definition content results from lexico-syntactic choices taken to meet human user needs. These needs can be linguistic, cognitive, and practical, depending on users' level of expertise, age, background, and the task being performed. Thus, the same defining content, for instance 'four legged', can be expressed as 'quadruped' when defining for specialists, and as 'that has four legs' when defining for laypersons.

Non-linguistic concretizations use symbols that also result from formalism choices adapted to the target audience and context of use (machine-readable logical formalisms and their different syntactic forms, first order logic, chemical symbols, etc.). The choice of formalism depends, for instance, on the formalism's syntactic and semantic properties, its expressivity, and the extent to which the formalism is known to the target audience.

A widely used formal language for representing ontologies is the Web Ontology Language (OWL). In OWL, 'class expressions' function in a manner analogous to the necessary conditions previously discussed in relation to logical forms of definitions (Section 2.2).¹⁰ OWL ontologies can have equivalence axioms which indicate that one or more class expressions are necessary and sufficient, as well as subclass axioms which indicate that a class expression represents a necessary condition. Together, OWL axioms serve to restrict the intended meaning of a term in an ontology by imposing conditions that must be satisfied in all models. In many of the ontologies where OWL has been used, it has proved difficult or impractical to provide statements of sufficient conditions; the axioms consist primarily of subclass axioms, i.e., they are partial definitions.

Whether in textual or logical form, definitions in ontologies are part of an ontology element, analogous to definitions in a dictionary where they are part of a dictionary entry. Other elements of an ontology likewise correspond to lexically relevant information¹¹, for example:

- term (label, preferred label/name/term, synonym, etc.)
- unique identifier (IRI),
- indication of domain or scope,
- note or comment,
- example,
- illustration (graphic images, photographs).

Figure 2 shows an example of information types associated with the *leukocyte* element (class) in the Cell Ontology.

¹⁰ See '9.1 Class Expressions', https://www.w3.org/TR/owl2-syntax/#Class_Expression.

¹¹ As they can be visualized, for example, in the BioPortal tool, <http://biportal.bioontology.org>. We only include lexically relevant information types.

The screenshot shows the BioPortal interface for the Cell Ontology (CL). The left sidebar contains a tree view of classes, with 'leukocyte' selected. The main content area displays the details for 'leukocyte' (CL:0000738). The details are organized into tabs: Details, Visualization, Notes (0), and Class Mappings (70). The 'Details' tab is active, showing various properties and their values.

Property	Value
Preferred Name	leukocyte
ID	http://purl.obolibrary.org/obo/CL_0000738
capable of	ameboidal-type cell migration
database_cross_reference	FMA:62852 BTO:0000751 CALOHA:TS-0549
definition	An achromatic cell of the myeloid or lymphoid lineages capable of ameboid movement, found in blood or other tissue.
develops_from	hematopoietic stem cell
has_exact_synonym	leucocyte white blood cell immune cell
has_obo_namespace	cell
label	leukocyte
part_of	immune system
prefixIRI	CL:0000738
prefLabel	leukocyte
treeView	immune system hematopoietic stem cell
subClassOf	motile cell hematopoietic cell

The footer of the page contains the following text: "The National Center for Biomedical Ontology is one of the National Centers for Biomedical Computing supported by the NHGRI, the NHLBI, and the NIH Common Fund under grant U54-HG004028. Copyright © 2005-2016, The Board of Trustees of Leland Stanford Junior University. All rights reserved. NCBO Website Release Notes Terms of Use Privacy Policy How to Cite"

Figure 2: Display illustrating the information types for *leukocyte* in the Cell Ontology (CL)¹²

In OWL, these are rendered as ‘annotation axioms’.¹³ These different types of information entity complement each other to provide users with a specification of the defined term and its referent. We will come back to this in Section 3, as this *complementarity principle* is important to understand the cognitive function of definitions.

2.4. DEF4: defining as an act of communication

Finally, ‘defining’ refers to a communication act with *sender* and *receiver* participating, which consists in the sender communicating to a receiver the definition contents (DEF2) by means of a representational artifact (DEF3). This process can be subdivided into different parts, including concretization on a communicable medium, realization by an oral speech act, and interpretation by an audience.

¹² Source: BioPortal,

https://bioportal.bioontology.org/ontologies/CL/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FCL_0000738.

¹³ See ‘10.2 Annotation Axioms’, http://www.w3.org/TR/owl2-syntax/#Annotation_Axioms.

Two major types of communication situation correspond to two subtypes of *defining act* and involve specific modalities, namely:

DEF4a: defining act *in presentia*

A defining act in which both sender and receiver are present and are communicating orally,¹⁴ for instance when an ontologist communicates with a subject-matter expert or a teacher with her students;

DEF4b: defining act *in absentia*

A defining act in which only the receiver is present, which consists in the receiver's consulting a definition produced earlier. These acts are, for example, situations in which didactic, scientific, and legal texts, or terminological, lexicographic, and encyclopedic dictionaries are consulted.

The defining act performed when an ontology or dictionary is consulted takes place in an *in absentia* communication situation: the sending of the definition content is spatiotemporally remote from its reception.

We will not further address these aspects here, as the main focus of our analysis of definitions in ontologies is on their content and form (DEF2 and DEF3), and on their functions.

3. Functions of definitions

We distinguish two main functions of definitions: cognitive and linguistic. These functions motivate the adaptation of the definition contents and form so as to be relevant for different target audiences and contexts of use.

3.1. Cognitive function of definitions

Definitions have primarily a cognitive function. That is, they produce some effect in the cognitive systems of their receivers. This function consists in reconfiguring and sometimes augmenting the receiver's beliefs in such a way as to fill the gap in knowledge that is implied by her definition consulting act (see also Sager 1990: 102). Indeed, definitions provide knowledge and beliefs about the objects, processes, and so forth that are essential to the everyday activities and practices of domain experts as reflected in their specialized vocabularies.

Definitions are consulted because it is presumed that the author of the definition has some knowledge about the term's meaning and properties of the term's referent that

¹⁴ Note that we exclude ostension or demonstration from the extension of the defining act in the sense of DEF4. This modality consists in designating a particular object or a particular situation, for example, by pointing to it. This kind of communication act is sometimes called 'ostensive definition' or 'demonstrative definition' (Sager 1990, Weinreich 1970). However, these acts are limited to perception and the cognitive function of definitions involves more than this (see Section 3.1).

the receiver lacks (Sager 1990: 101-102, 112).¹⁵ Any such definition consulting act should therefore involve some modification in the body of knowledge and beliefs of the receiver, including the sort of modification that consists in adding greater confidence to the receiver's beliefs.

To see why this is so, let us go back to the complementarity principle introduced in Section 2.3.2. If the receiver consults an ontology or a dictionary to look up a definition, then she presumably already possesses complementary information type(s), such as a term or some perceptual knowledge of its referent, as shown in Figure 3. Therefore, the receiver's need lies at the semantic level of knowledge and beliefs.

What you know	What you are presumably looking for		
	ligament	<i>Dense regular connective tissue connecting two or more adjacent skeletal elements or supporting an organ.</i>	
ligament		x	x
<i>Dense regular connective tissue connecting two or more adjacent skeletal elements or supporting an organ.</i>	x		x
	x	x	

Figure 3: Complementarity of information types corresponding to three states of knowledge: of term, of definition, of referent (or of an image of the referent)

The need on the part of the receiver for the definitional content can be either real or only pretended, as for example when a teacher pretends not to know the meanings of words when asking questions of her students. In the communication situations associated with standard uses of ontologies and dictionaries this need is presupposed. The definer is not directly acquainted with the receivers; but she can nonetheless be assured that the typical users of the artifacts she produces are in need of cognitive reconfiguration or augmentation of just this sort through the enhancement of her lexical competences.

3.1.1. Theory of lexical competence

To put the above in terms of lexical competence, consulting an ontology or dictionary resource implies either (i) that the receiver's lexical competence is lacking (perhaps because it is marked by uncertainty and thus in need of confirmation), or (ii) that the receiver's lexical competence is to be modified to fit a specific context of use.

¹⁵ That is, some knowledge about the intended meaning of a term by a group of competent speakers to refer to something that is the object of the definition.

Marconi (1997: 2) divides lexical competence into two independent and cognitively motivated competences, one *inferential* (IC) and one *referential* (RC), further subdivided into *naming* and *application* abilities (NRC and ARC). These correspond, respectively, to a speaker's ability to:

IC: “have access to a network of connections between [a] word and other words and linguistic expressions”, and thus perform different types of inferences;

NRC: name objects and circumstances in the world; that is, select “the right word in response to a given object or circumstance” (*naming*);

ARC: apply a word to objects and circumstances in the world; that is, select “the right object or circumstance in response to a given word” (*application*).

These competences rely on two distinct systems:¹⁶ the former (IC) on a semantic system; the latter (RC) on the perceptual and motor system. Neuropsychological studies show that each of these abilities can be lost or seriously damaged while the other remains somewhat intact (Riddoch and Humphreys 1987b). Both systems nevertheless interact and inferential competence plays a role in many referential performances, just as referential competences can enrich inferential competences.

In this light we propose that ontology elements and dictionary entries *adjust* receivers' lexical competence to converge towards that of competent speakers of a given domain (e.g., microbiology) in a given context of use (e.g., data annotation with an ontology of microbiology).

Ontologies and dictionaries always have this cognitive function of lexical competence adjustment. Most if not all of the contents of the different fields of a dictionary entry contribute to the realization of this function, and something similar is true also in the case of ontologies (Figure 5 in Section 3.1.2).

To understand the specific cognitive function of definitions, we go back to Marconi's theory of lexical competence in (1997: 70-73), which puts forward the further hypothesis that “inferential competence may include several conceptually distinct and mentally separate abilities.” He mentions, for instance, two specific sub-competences involving access to the *output lexicon* and the *semantic lexicon*.

– Output lexicon: “the words themselves (in either their phonological or graphic format)”.

– Semantic lexicon: concepts or semantic representations “accessible from both words and pictures” and providing access to corresponding entities in the world.

We call the former ‘output inferential competence’ and the latter, ‘semantic inferential competence’.¹⁷ We define the output inferential competence as the ability, not

¹⁶ See Marconi (1997: 61-64, 141-142).

¹⁷ Not to be confused with the semantic competence that is part of our general linguistic competence (Marconi 1997: 77)

only to access the semantic lexicon, but also to make connections between word forms and sounds independently of their semantics. For example, to play word games where you go from ‘ligament’ to ‘lineament’ and ‘liniment’ (which have no connected meanings). We define the semantic inferential competence as the ability, not only to access the output lexicon and the world, but also to make inferences and connections between meanings, descriptions, definitions (independently of lexicalizations or of the world). For example, if someone tells you that they tore a ligament and you know the definition of that term, you might infer that the affected anatomic parts that should be held together are not anymore and as a result there would be considerable pain. This subdivision is supported by cases of individuals capable of describing the properties and functions of an object without being able to name it (Warrington 1985: 341-342; Riddoch and Humphreys: 1987a, 132; Shallice 1988: 292 ff.) McCarthy and Warrington 1988: 429). While Marconi does not pursue this distinction further, he nevertheless notes that “the ability to define words (word to definition) may be dissociated from the ability to find the word corresponding to a given verbal definition or to a description of the word’s referent (definition to word).” (*op. cit.*)

The distinction is relevant for understanding what kind of cognitive mechanism or process lies behind the act of defining: defining (DEF4) is an act that goes directly or indirectly from the word to the definition. We schematize these distinctions in Figure 4.

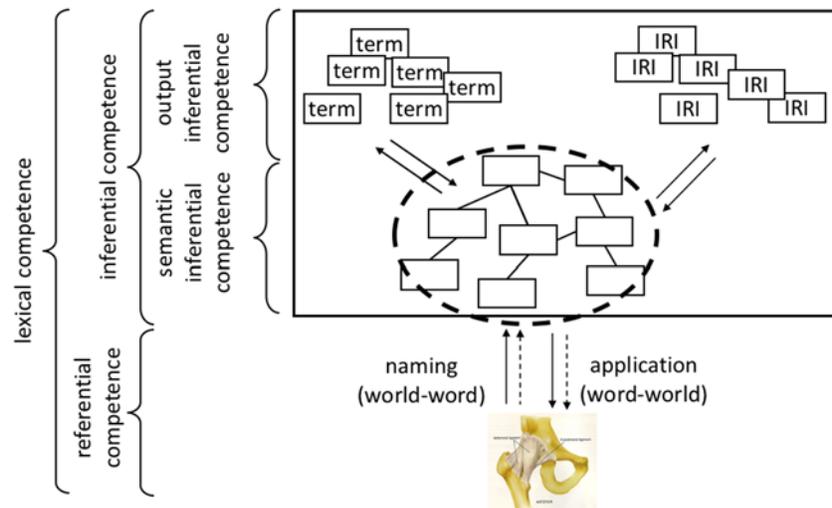


Figure 4: Lexical competence and its sub-competences (based on Marconi 1995, 1997)

Following these subdivisions, we propose that a definition adjusts the overall lexical competence of its receivers by adjusting their inferential competence, and, more specifically, their semantic inferential competence. The effect on referential competence is realized indirectly, for example, through those words in the definition in relation to which the receiver already enjoys referential competence.

3.1.2. Complementary information types for complementary adjustments

Based on Marconi’s intuition “that the two sides of lexical competence, inferential and referential, mostly rely upon different kinds of information” (1995: 149) and considering the above subdivision of inferential competence, each information type in a dictionary entry and ontology element can be paired to a sub-type of competence.

To see this, we associate the three types of lexical sub-competences introduced above,¹⁸ namely:

- referential competence,
- semantic inferential competence,
- output inferential competence,

with the complementary lexically relevant information types in an ontology element presented in Section 2.3.2. These pairings are illustrated in Figure 5.

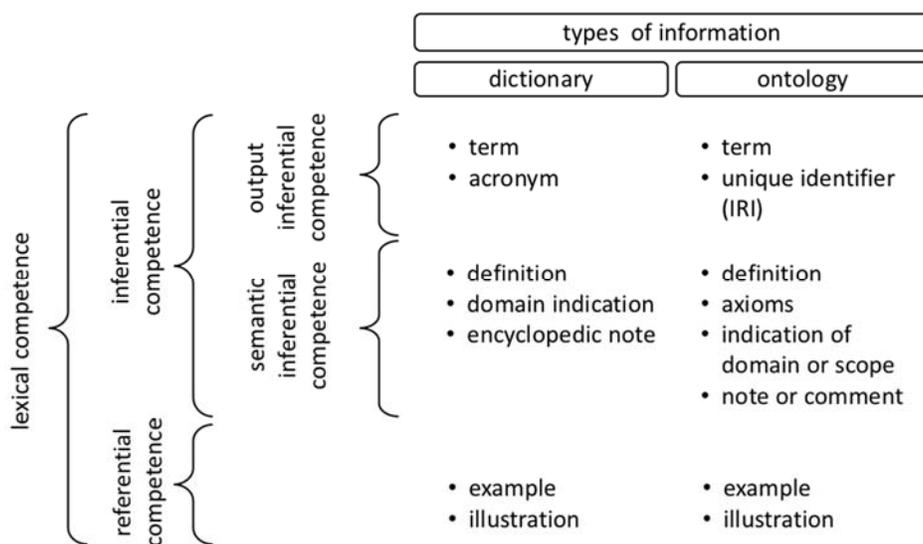


Figure 5: Pairings of information types in a dictionary entry and ontology element with the sub-competences in which they are involved

– Referential competence: Ontologies can include *examples* among their elements; and also, though more rarely, *illustrations* (for example, photographic images). Both allow users to recognize a described portion of reality on the basis of exposure to information types related to referential competence. In the case of examples, the example text involved will standardly call in aid only the user’s referential competence — thus it will not explicitly convey any information pertaining to the properties of the referent of the

¹⁸ This parallel between different types of competences and different types of dictionaries is addressed several times by Marconi (1997: 56, 66-67, 114, 146, 157). In Seppälä (2012), we extended it to dictionary entry fields. Here we apply the proposed pairings to ontologies.

sort that would be provided by a definition. Illustrations can be provided by either the ontology developers in the form of links to images accessible through the ontology interface,¹⁹ or by the ontology's users, through the annotation of images with ontology classes – for example, when images of plant formations are annotated with terms from the Plant Ontology (PO) (Lingutla, *et al.* 2014).

– Semantic inferential competence: Those information types in an ontology that relate to semantic inferential competences are the *textual definitions* and *axioms*, the *notes* (in 'comment' annotation properties), and *indications of the domain* or *scope* of the ontology. These parts of an ontology element provide information that contributes to a user's understanding of the intended meaning of the ontology classes and their use in inferential, i.e., logical, operations. They convey the body of knowledge and beliefs that can reasonably be considered to be shared by both ontologists and domain experts when they use a given term.

– Output inferential competence: The information types that relate to this competence are the label (a word or a phrase), including not only preferred label and synonyms but also the IRIs identifying classes or relations. IRIs are included under this heading insofar as ontologies are manipulated by machines and the IRIs are the symbolic forms that a machine uses.

Thus, if we consider that each lexically relevant information type in an ontology element and dictionary entry is involved in one of the three lexical sub-competences (referential, output inferential, semantic inferential), then we can say that definitions have primarily the function of adjusting receiver's semantic inferential competence, which is concerned with beliefs and reasoning.

The cognitive function of definitions, in sum, is to bring about a belief reconfiguration or belief augmentation on the part of the receiver such that it allows her to competently use the defined term in inferential processes and, indirectly, also in referential ones. For this adjustment to be successful, the definition has to be relevant with respect to a specific target audience and context of use. We address definition relevance in the context of ontologies in the next section in relation to the linguistic function of definitions.

3.2. Linguistic function of definitions

In most cases, consulting an ontology or dictionary reflects, among other things, a need to align oneself with a certain pre-existing lexical use. Even though consulting such resources does not necessarily answer a lexical question as such, the fact that both ontologies and dictionaries include lexical units implies, at least indirectly, that they fulfill a linguistic function.

¹⁹ See the Foundational Model of Anatomy (FMA) ontology browser at <http://xiphoid.biostr.washington.edu/fma/index.html>.

Someone consulting an ontology or dictionary, and thus a definition, aims to attune their lexical competence in order to promote their use of given linguistic signs in a way that converges towards that of competent speakers.²⁰ In ontologies and terminological dictionaries, these competent speakers are either (i) domain experts, where the resource in question covers domain-specific terms, or (ii) ontology or terminology experts, for matters pertaining to domain-independent ontology or terminology terms.

A single lexical unit (graphic or phonic sign) can have different semantic values depending on the context of use, for example, on the domain of expertise (e.g., banking, nephrology) and the task at hand (e.g., understanding a text, annotating scientific data). Semantic value is thus determined in part by the context of use of given competent speakers. Definitions adjust receivers' lexical competence toward the corresponding global norm of use, that is, a semantic value acknowledged by a given speaker community.

Yet definitions are also often used stipulatively (Ajdukiewicz 1974, Gupta 2015), to describe or prescribe a use that is not the global norm. The effect of the definer creating the definition is to establish local convergences (between themselves and successive readers) to avoid confusion with and ambiguities within the global convergent norm of use.

Both global and local uses can appear in free discourse and text that constitute the basis for the defining activity encapsulated in ontologies and dictionaries. Ontologists and dictionary authors are, after all, mediators — they usually do not create norms of use but rather mediate between, for instance, authors of scientific textbooks and the users of terminologies who require support from definitions. Their secondary defining activity consists in making explicit the meanings underlying uses of terms of distinct provenance: either a global norm of use or a local use.

Once a definition is included in an ontology or dictionary, it thenceforth expresses a lexical use, which can be more or less stipulative (prescriptive) depending on the context of use of the ontology or dictionary, the intention of the definer, the attitude of the receiver, and so on. A definition can thus be regarded as being on a scale that ranges from describing a use (descriptive definition) to having a regulatory or normative function on use (stipulative or prescriptive definition).

Although an empirical question, it is reasonable to assume that the more a definition is descriptive and aimed at conveying the meaning of a term, for example, for understanding a text, the more it is likely that it will have a (proto)typical logical form, with a combination of necessary and (proto)typical features. On the contrary, the more a definition aims at stipulating a meaning, the more it tends to include only necessary (and possibly sufficient) conditions, which allow users to disambiguate the defined term.

²⁰ Throughout this section, we use Marconi's theory of Lexical Competence (Marconi 1997). For a similar pragmatic approach to competence adjustment and to its directionality (influenced by John Searle), see (Riegel 1990: 100-101).

The relationships between types of resources, their lexical uses, and the logical forms, utility, and cognitive effect of their definitions is illustrated in Table 1.

	<i>Type of resource</i>	
	Lexicographic, lexico-semantic, terminological	Ontologies, standards, normative terminologies
<i>Lexical use</i>	describe	prescribe
<i>Logical form of definitions</i>	(proto)typical	necessary & sufficient conditions
<i>Utility</i>	conveying meaning	disambiguating meaning
<i>Cognitive effect</i>	knowledge/belief augmenting	knowledge/belief reconfiguring

Table 1: Relationships between types of resources, their lexical uses, and the logical forms, utility, and cognitive effect of their definitions

The table shows the spectrum of properties of more lexically-oriented resources in the middle column and ontologically-oriented resources on the right. The former tend to have more descriptive and proto-typical definitions, employed mostly to convey meaning. The latter tend to include more prescriptive definitions with necessary and, possibly, sufficient conditions, employed mostly to disambiguate meanings. The corresponding cognitive effects tend to be, respectively, more belief-augmentating and belief-reconfiguring. Note that whenever a definition fulfills a linguistic function a cognitive function is also fulfilled — but that the reverse does not hold.

In sum, the linguistic function of a definition is to convey or disambiguate the semantic value of a term in a more or less descriptive or stipulative way, by delimiting its intension and extension by means of more or less proto)typical and classical definitions. The resulting cognitive effects on the body of knowledge and beliefs of the receiver are more or less to augment and reconfigure them.

4. Functions of definitions in ontologies

4.1. Stipulation of the intended meaning of a term

In ontologies, the represented knowledge is not always related to a lexical unit that is naturally used by speakers. The represented classes are usually labeled with terms from a controlled vocabulary with an intended meaning, where terms from the vocabulary may or may not be used in natural-language contexts. Whether an ontology term is commonly used by domain experts, its definition specifies the intended meaning of the ontology terms in a normative way. Definitions in ontologies allow ontology users to use these terms in a competent manner; ontologies enhance users' lexical competence with respect to ontology terms by providing definitions that clarify and disambiguate.

The corresponding adjustment of semantic inferential competences is, therefore, mostly a belief-reconfiguring one — as opposed to a belief-augmenting adjustment. When consulting a definition in an ontology, the receiver (usually a domain expert) has to set aside the body of knowledge accumulated over the years about the terms of her domain and their referents, and restrict the meaning of the term to only those items of knowledge intended by the ontology developers. This involves some reconfiguration in her beliefs.

4.2. Term disambiguation

Ontologies and controlled vocabularies aim at aligning the lexical use of their users to achieve intra- and inter-personal consistency, for example, when annotating scientific data or integrating databases with an ontology. Correspondingly, definitions fulfill a stipulative linguistic function; that of adjusting the receiver's lexical competence towards a local use. In order for a definition to realize this function, it has to be tailored in such a way that its logical form leaves no room for ambiguity. It therefore has a disambiguating function.

To realize this function, ontologies would ideally provide Aristotelian definitions in which the genus (more precisely, the immediate superordinate category or *genus proximus*) is specified along with the differentiae to provide a statement of individually necessary and jointly sufficient conditions that precisely distinguishes the intended meaning of the term from that of neighboring terms. Provision of such Aristotelian definitions is costly, however, and in some cases it is not possible at all because of lack of scientific knowledge; thus ontologies (especially large ontologies) often make do merely with the statement of necessary conditions as tool for disambiguation.

Where definitions are provided using a formal language such as OWL, they take the form of axioms which serve to disambiguate terms in a way that is analogous to the way textual definitions serve this purpose. Every subclass axiom represents a necessary condition that all instances in the extension of the term need to satisfy. These axioms serve to determine the extension of a term by restricting it to those entities meeting the asserted condition. Each additional axiom restricts the extension further, though it is in

many cases not possible to restrict the term to only its intended extension by providing conditions that are jointly sufficient. When a classical definition is possible, an equivalence axiom is used as shown in the axiom defining the IDO term *bacteremia*.

```
bacteremia
EquivalentTo
  infection and
  (has_part some
    (infectious agent and Bacteria and
      (located_in some blood)))
```

For the most part, a class (as opposed to a class expression) serving as relatum in a subclass and equivalence axiom should correspond directly to the genus in the textual definition as in the case of ‘achromatic cell’ in the definition of *leukocyte* (Figure 1). The other defining conditions are expressed by non-atomic class expressions.

```
unblinding process
SubClassOf
  planned process
  (part_of some study design execution)
  and (part_of some informing subject of study arm)
```

Here, two axioms define the term *unblinding process* in the Ontology for Biomedical Investigations (OBI). ‘Planned process’ is the asserted superclass of ‘unblinding process’, as in: ‘unblinding process is_a planned process’. ‘Part_of some study design execution’ and ‘part_of some informing subject of study arm’ are class expressions that in conjunction restrict the extension of ‘unbinding process’ to only those planned processes that are part of a study design execution and that inform the subjects about the study arm in which they participate.

The logic of the ‘subClassOf’ relation is that the members of the defined class are also members of all the other classes specified in the conjoined class expressions. Thus, all unblinding processes are members of the class ‘planned process’, as well as of the classes of things that are part of a study design execution and things that inform the subjects about the study arm in which they participate.

4.3. Functions of logical definitions

In addition to the cognitive and linguistic functions described above, which apply both to textual and logical definitions, we distinguish three primary functions of logical definitions: instance classification and consistency checking, taxonomic schematization, and regularizing expression of facts.

4.3.1 Instance classification and consistency checking

Classical definitions function in instance classification and consistency checking. Necessary conditions serve as checklists for determining whether an instance is consistent

with the classes of which it is asserted to be a member. When an instance's properties are consistent with sufficient conditions for a class, that instance can be asserted to be of that class. Indeed, the linguistic function of a definition is to convey the semantic value of a term by delimiting its intension and extension. Definitions in ontologies include statements of necessary conditions and thus allow us to define intensions of their terms and thereby enable ontologies to be used as reliable heuristics for such classification tasks.

This function is further useful for identifying errors in definitions, since it allows the definer to test the scope of a definition by seeing whether it classifies the right instances and is able to exclude unwanted ones as inconsistent. The definition content may thus be checked and corrected to ensure that it is about the definition object and only the definition object. For example, if there are exemplars of the definition object that don't concord with the definition then the defining content is adjusted so as to include those. Similarly, if there are entities that concord with the definition, but are not exemplars of the intended definition object, then the defining content is adjusted to exclude them.

4.3.2. Taxonomic schematization

We call the second function 'taxonomic schematization'. When employed in this capacity, the logical definition of a class provides a schema or template for the axioms of its subclasses. The goal is to provide robust, principled taxonomic relations between parent, child, and sibling classes. The axioms specified for each class are true of all its subclasses. This makes it possible to use axioms to specify differentiae for its child classes, in other words, to use these axioms as templates for the axioms of the subclasses, as well as for the contents of the associated textual definitions (Seppälä 2012, 2015).

This can be done by asserting a relational axiom for the parent class relating it to some other kind of entity (e.g., by writing an axiom for a class X asserting that any X *part_of* some Y). For every subclass of X, a subclass of Y can then be distinguished. The property relating both relata can also be a sub-property of the more general one. For example, the axiom specifying the term *infection* in the Infectious Disease Ontology (IDO):

```
infection
SubClassOf
  material entity
  (has_part some infectious agent) and
  (part_of some extended organism)
```

can be used to generate the axioms of its child terms, such as *bacteremia* (see Figure 6). Here the class expression 'infectious agent and Bacteria' is a subclass of 'infectious agent' and so further specifies what kind of 'infectious agent' characterizes bacteremia.

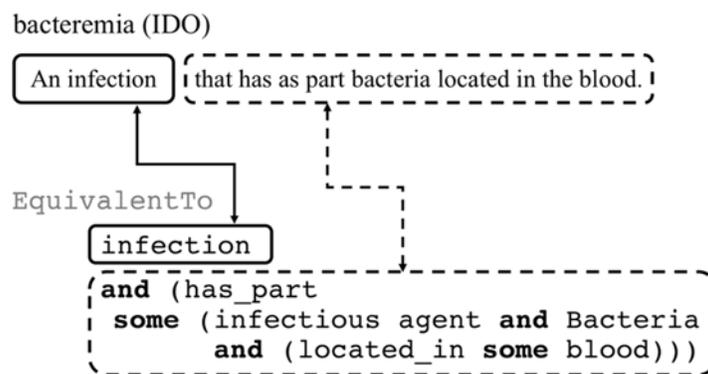


Figure 6: Correspondences in the parts of the textual definition and the axioms of the IDO term *bacteremia*

4.3.3. Regularizing expression of facts

An ontology can be considered a specification of a controlled vocabulary for expressing facts in a given domain. Such a vocabulary is much sparser than the vocabulary that would be used to express these facts in natural language, that is, there is a one-many correspondence between ontology terms and words in domain-relevant portions of natural language. This means that the syntax for expressing facts (i.e., assertions between instances) using ontology terms necessarily diverges from the syntax used for expressing the same facts in natural language.

An important function of axioms in ontologies is to provide a schematic indication of how this should be done. Thus, axioms complement textual definitions in contributing cognitively towards regularizing users' use of terms. Consider, for example, the class expression '*develops_from* some hematopoietic stem cell' as it occurs in one of the axioms involving the term 'leukocyte' in the Cell Type Ontology (CL) (see Sections 2.2 and 2.3.1). In this expression, the relation '*of the ... lineages*' in natural language is expressed at the logical level by the '*develops_from*' relation that is part of the controlled vocabulary of the ontology.

4.4. Synthesis

Textual and logical definitions in ontologies have overlapping and complementary functions. Both adjust receivers' (humans' and machines') lexical competence by describing or stipulating a term's intended meaning. The particular context of use of ontologies gives precedence to the logical definition, since it is the one used by the reasoner to perform logical operations. Therefore, the contents of the definition have to be adapted for use by machines. One such adaptation is to include necessary, and whenever possible sufficient, conditions.

Yet, stating only necessary (and sufficient) conditions in a textual definition might be too limited for an adequate understanding of the defined term. It may be useful to add extra information about the defined term's referent, such as typical features. But such

non-necessary conditions could, in logics such as description logics, make reasoning more difficult. Thus, any extra information that might be useful for human understanding should be included, for such logics, in the form of a comment that may be complemented with examples. As we saw in Sections 2.3.2 and 3.1, all these information types – definitions, comments, notes, examples, etc., – complement each other to enhance an ontology user’s overall lexical competence; the definition adjusts it to a specific context of use, which in ontologies requires disambiguating terms. In that sense, the textual definition in the ‘comment’ or ‘definition’ annotation property contains the information that is relevant in the context of ontologies and their uses. The rest of the information could also be relevant, i.e. potentially defining, for example in a terminological dictionary context, but not in an ontology. The complementary information must however be controlled to avoid introducing ambiguity.

Furthermore, in order to ensure consistency in ontology development and use, the textual and logical definitions of a term must convey the same type of content. Thus, axioms can be used as content templates for lower-level categories and for textual definitions (provided the axioms are complete, see Section 5 below). The controlled vocabulary used in axioms allows the user to check the intended meaning of a natural language expression in the textual definition. However, as noted by Stevens, *et al.* (2011: 6), textual definitions are not simple sentence-by-sentence verbalizations of axioms. The axioms create unordered lists of sentences or sentence fragments that can involve redundancies. A textual definition is rather a grouping of one or more axioms that form a non-redundant, fluent paragraph. To illustrate this, see the difference in the order of the differentiae and the corresponding axioms in the textual and logical definitions of the IDO term *amebiasis* (Figure 8 in Section 5.2 below). The expressions used in natural language definitions are more idiomatic. Expressions such as ‘continuant_part_of’ or ‘inheres_in’ are after all not very natural.

In this respect, an ontology element functions like a dictionary entry in providing complementary information types that participate in the adjustment of the different sub-competences of the ontology users’ overall lexical competence.

5. Similarities and differences between textual and logical definitions in ontologies

The linguistic function of textual definitions in ontologies is to specify the intended meaning of the ontology terms in order to avoid ambiguities and errors when for example annotating biomedical research texts and importing terms into other ontologies. Of course, this is also the function performed by logical definitions, as we saw in the previous section. However, not all users of ontologies are willing to work with logical definitions, and even the creators of logical definitions may require a textual definition as starting point.

Ontologies provide textual and logical definitions to enhance human users’ and computational systems’ lexical competence. They adjust these receivers’ semantic inferential competences by reconfiguring their body of knowledge associated with the use

of given terms. The goal of textual and logical definitions in ontologies is to align their use of a term with that of the rest of the community of competent speakers using the same ontology. Therefore, as we saw, definitions must be relevant for that purpose, that is, provide the relevant kind of content (DEF2) in an adequate unambiguous form (DEF3).

Logical and textual definitions differ in form, but are of course not completely distinct. In the ideal case, indeed, they will be logically equivalent. At a minimum, certain correspondences will exist between the phrases used in textual and in the corresponding logical definitions (Stevens, *et al.* 2011: 3). These correspondences can be used to provide guidelines for identifying problems with definition contents as well as indications for quality assurance of definitions, and thus help developers improve their ontologies (Seppälä, *et al.* 2014).²¹

In this last section, we compare textual and logical definitions in ontologies to reveal similarities and differences between both forms, and show what kinds of issues or inconsistencies can be identified through these comparisons. By working through examples of the correspondence between parts of textual and logical definitions, we show how to compare and contrast each, and how each perspective reveals areas for improvement. We identify at least four types of correspondences.

5.1. Exact correspondence

Figure 7 shows that the parts of the textual definition of *dead-end host* in IDO correspond exactly to the logical definition by necessary and sufficient conditions (indicated by the ‘EquivalentTo’ property). The only difference is in the natural language expression (‘bearing’) that is used for the ‘has_role’ ontological relation — perhaps to avoid the seemingly redundant use of ‘role’ twice in the textual definition. Here, the logical part is useful to clarify the intended meaning of the natural language expression.

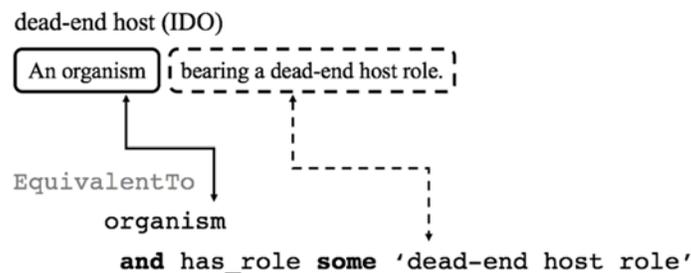


Figure 7: Correspondences in the parts of the textual definition and the axioms of the IDO term *dead-end host*

²¹ We can also apply the sorts of semi-automated methods being used in the terminological world (Seppälä 2012, 2015) both in establishing such correspondences and in identifying quality issues in the textual definitions that might point to quality issues in their logical counterparts.

5.2. Structural correspondence but more specific content in textual definitions than in axioms

Figure 8 shows that both differentiae in the textual definition of the IDO term *amebiasis* contain information of the type expressed in the subclass axioms inherited from the parent class *infection* (see dashed and dotted lines). However, part of the content conveyed by the textual definition of *amebiasis* is more specific than the expressions in the axiom: ‘colon’ and ‘organism of the Species *Entamoeba histolytica*’ are respectively subclasses of the classes ‘part_of extended organism’ and ‘infectious agent’ in the axiom.

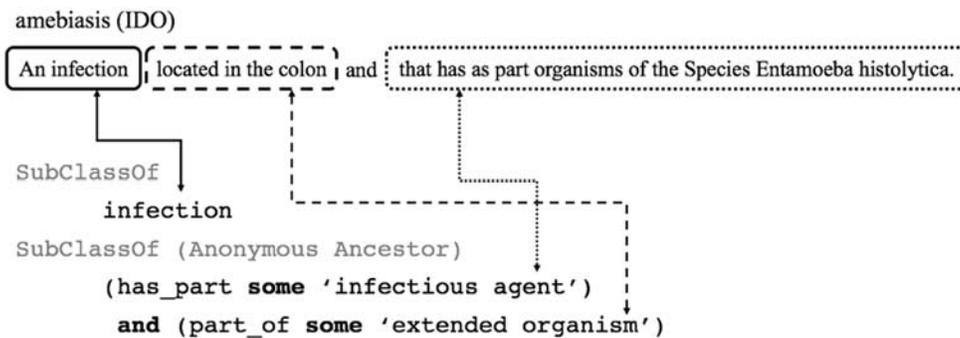


Figure 8: Correspondences in the parts of the textual definition and the axioms of the IDO term *amebiasis*

If the axioms of a class are relevant for distinguishing its subclasses, then the set of axioms should be used as a template for the subclasses—see the taxonomic schematization function of logical definitions in Section 3.3.2. The set of axioms of each subclass then instantiates the template by replacing the same set of elements with classes of similar specificity. For example, if the differentiae in the class template include conditions on 'has part' and 'part of', all the subclasses of that class should use specializations of those conditions in their differentiae. Using this principle, we can check the adequacy of definition contents by identifying mismatches in the levels of specificity of subclass axioms. For example, if part of a term’s logical definition matches an expression inherited from a superclass, it might be a sign that the logical definition of the defined subclass is missing a more specific subclass axiom to distinguish it from its parent and neighboring classes. If this is the case, the textual definition can be used as a basis for creating this more specific axiom. For instance, in the *amebiasis* example of Figure 8, the class expression ‘part_of some extended organism’ in the definition of the term’s superordinate *infection*, could be specialized with the following expression:²²

located_in **some** colon.

²² According to the definition of ‘extended organism’, anything that is located in an organism (e.g., in the interior of the colon) is part of the extended organism (e.g., the body) of which that organism is a part (a colon is part of a body).

This more specific expression could be added to the logical definition of *amebiasis* if the ontology provides the corresponding terms (class and object property).

5.3. Missing axioms

The term *transmission process* from IDO has the textual definition *A process that is the means during which the pathogen is transmitted directly or indirectly from its natural reservoir, a susceptible host or source to a new host.* However, the only logical axiom given is that it is a subclass of ‘process’. The textual differentiae parts have no correspondence in the logical axiom, for example:

- that there are participants including pathogen and susceptible host;
- that the process occurs in part in a natural reservoir or infected source and in part in the host.

Here too, the textual parts can be used for creating the corresponding axioms.

5.4. Redundant parts of axioms or definitions

Logical parts may contain axioms defining other terms. Figure 9 shows that parts of the axioms defining *antiseptic role* in IDO imply two sorts of redundancies: (i) the first axiom (see dashed lines) includes the logical definition (‘EquivalentTo’) of the class ‘antimicrobial’, i.e., a material entity that has an antimicrobial disposition; (ii) the second axiom includes the subclass axioms defining the term *antimicrobial disposition* (see grey lines).

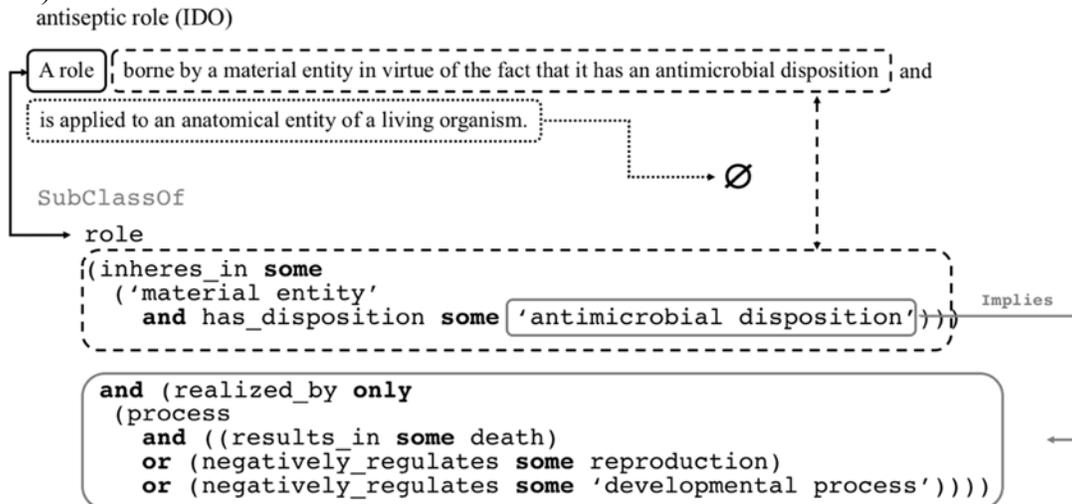


Figure 9: Redundant axioms in the logical definition of the IDO term *antiseptic role*

These redundancies should not be a problem at the logical level, since the inferences that are made based on the logical expressions end up being the same computationally. However, at the textual level, this amounts to defining another term within the definition

of the defined term. This lacks conciseness and is generally considered bad practice.²³ It unnecessarily overloads the contents of the definition — imagine if each term of a definition were replaced by its definition. More importantly, the reader might not recognize that it is the definition of another term and fail to link the defined term with that other one. Whenever a textual definition contains the definition of another term from the same ontology or an imported ontology, this sub-definition should be replaced by the corresponding term. If the reader does not know the term used in the definition, she can (in principle) look it up in the ontology.

Going back to Figure 9, the redundancy in the second part also reveals an error in the logical definition. The first clause says that everything in which an antiseptic role inheres also has an antimicrobial disposition. The second clause gives conditions on the realization of the antiseptic role, which are taken again from the definition of ‘antimicrobial disposition.’ So, as it appears in the logical definition, antiseptic role is stated to be a subclass of antimicrobial disposition. However, antiseptic roles are different from antimicrobial dispositions because antiseptics are used for a purpose (‘is applied to’) but they have their action because they have an antimicrobial disposition. As roles and dispositions are disjoint, ‘antiseptic role’ cannot be a subclass of ‘antimicrobial disposition’.

The problem is that in order for the role to be differentiated from the disposition, the realization needs to be different from that of the disposition. The realization of the antiseptic role *also* includes the application of the material, for example, of an antiseptic liquid on the skin. The dotted lines in the textual definition (pointing to the \emptyset symbol) show that this information is included in the textual definition, but is missing from the axioms.

The definition can be reformulated to fix these errors and, in the process, remove the repeated parts of the definition as follows:

```
antiseptic role
SubClassOf
  'role' and
  inheres_in some 'antimicrobial'
  (realized_in only
    'planned process' and
    (has_part
      ((realizes some 'antiseptic disposition') and
        (occurs_in some 'extended organism'))))
```

6. Conclusion

We examined the characteristics of the four types of things to which the term ‘definition’ refers: a *cognitive activity* (DEF1) that produces a *cognitive representation*

²³ See, for example, ISO 704 (2009: 28).

(DEF2), which constitutes the content of a *representational artifact* (DEF3) that concretizes this content and that is communicated in a *communication act* (DEF4).

We then focused on the cognitive and linguistic functions of definitions in general, and on the specific functions of textual and logical definitions in ontologies. We based our analysis of the cognitive and linguistic functions of definitions on a theory of lexical competence proposed by Marconi. Following the subdivisions of lexical competence suggested by Marconi and extending them with two more specific inferential competences, we put forward the following explanation: definitions adjust receivers' overall lexical competence by adjusting their inferential competence, and, more specifically, their semantic inferential competence. In ontologies, definitions bring about a reconfiguration of the receivers' existing body of knowledge regarding the intended referent of the defined term. Definitions and axioms adjust the knowledge or beliefs of the receiver in such a way that they become more closely aligned to the knowledge or beliefs of relevant experts.

We showed that to achieve an adequate adjustment, definitions must be adapted to allow the receivers to use a term in a way that converges toward the use of that term by the community of experts for a specific task. Therefore, linguistically speaking, definitions function to delimit the intended meaning of ontology terms to avoid ambiguities and errors when applied to such tasks as annotating scientific texts with ontology terms and reusing existing terms in new ontologies. To do so, definitions must be constructed in a way that is relevant to their context of use and their target audience.

In the context of use of ontologies, logical definitions are mainly intended for automatic reasoning systems; their symbolic and logical forms must be computer-tractable. Logical definitions take the form of axioms stating necessary and, whenever possible sufficient, conditions to delimit the intension and extension of a class term. To ensure a consistent use of the ontology terms by human users, their textual definitions should contain the same type of content as the corresponding logical definitions. Nevertheless, in virtue of the complementarity principle of information types, textual definitions can be complemented with notes and other useful comments included in the other parts of an ontology element.

In the last section, we analyzed four types of correspondences between textual and logical definitions to show what kinds of issues or inconsistencies can be identified by these comparisons. We suggested that comparing the parts of both forms of definitions may be useful for improving the quality of definitions in ontologies.

Acknowledgment

Work on this paper was supported in part by the Swiss National Science Foundation (SNSF) and in part by the National Institutes of Health (R01GM080646). We also thank Peter Elkin of the University at Buffalo for invaluable assistance.

Authors

Selja SEPPÄLÄ, University of Florida, sseppala@ufl.edu
Alan RUTTENBERG, University at Buffalo, alanruttenberg@gmail.com
Yonatan SCHREIBER, University at Buffalo, yonatan.schreiber@gmail.com
Barry SMITH, University at Buffalo, phismith@buffalo.edu

References

- AJDUKIEWICZ Kazimierz (1974): “Definitions”, *Pragmatic Logic*, Springer, Netherlands, p. 57-84.
- ARP Robert, SMITH Barry and SPEAR Andrew D. (2015): *Building Ontologies with Basic Formal Ontology*, Cambridge, MA, MIT Press.
- BODENREIDER Olivier and STEVENS Robert (2006): “Bio-ontologies: Current trends and future directions”, *Briefings in Bioinformatics*, 7, p. 256-274.
- BUCCELLA Agustina, CECHICH Alejandra, GENDARMI Domenico, LANUBILE Filippo, SEMERARO Giovanni and COLAGROSSI Attilio (2011) “Building a global normalized ontology for integrating geographic data sources”, *Computers & Geosciences*, 37, p. 893-916.
- CEUSTERS Werner and SMITH Barry (2010): “Foundations for a realist ontology of mental disease”, *Journal of Biomedical Semantics*, 1, p. 10.
- CHAURAND Jacques and MAZIÈRE Francine Eds (1990): *La définition*, Paris, Librairie Larousse.
- FONSECA Frederico T., EGENHOFER Max J., AGOURIS Peggy and CÂMARA Gilberto (2002): “Using ontologies for Integrated Geographic Information Systems”, *Transactions in GIS*, 6, p. 231-257.
- GUPTA Anil (2015): “Definitions”, *The Stanford Encyclopedia of Philosophy*, ZALTA E. N., Summer 2015 edition.
- HOEHNDORF Robert, SCHOFIELD Paul N. and GKOUTOS Georgios V. (2015): “The role of ontologies in biological and biomedical research: a functional perspective”, *Briefings in Bioinformatics*, 16, p. 1069.
- ISO 704 (2009): *Terminology work — Principles and methods (ISO 704:2009)*, Geneva, ISO.
- KONTOPOULOS Efstratios, MARTINOPOULOS Georgios, LAZAROU Despina and BASSILIADES Nick (2016): “An Ontology-Based Decision Support Tool for Optimizing Domestic Solar Hot Water System Selection”, *Journal of Cleaner Production*, 112, Part 5, p. 4636-4646.
- LINGUTLA Nikhil Tej, PREECE Justin, TODOROVIC Sinisa, COOPER Laurel, MOORE Laura and JAISWAL Pankaj (2014): “AISO: Annotation of Image Segments with Ontologies”, *Journal of Biomedical Semantics*, 5, p. 50.
- MARCONI Diego (1995): “On the Structure of Lexical Competence”, *Proceedings of the Aristotelian Society*, 95, p. 131-150.
- (1997): *Lexical Competence*, Cambridge, Massachusetts and London, England, The MIT Press.
- MARTIN Robert (1990): “La définition “naturelle””, *La définition*, Eds CHAURAND J. and MAZIÈRE F., Paris, Librairie Larousse, p. 86-95.

- MCCARTHY Rosaleen A. and WARRINGTON Elizabeth K. (1988): “Evidence for Modality-Specific Meaning Systems in the Brain”, *Nature*, 334, p. 428-430.
- PAVEL Silvia (2012): *The Pavel Terminology Tutorial*, <http://www.bt-tb.tpsgc-pwgsc.gc.ca/btb.php?lang=eng&cont=308>, (accessed 17/12/2015).
- PAVEL Silvia and NOLET Diane (2001): *Handbook of Terminology*, Canada, Public Works and Government Services - Translation Bureau.
- REBEYROLLE Josette (2000): *Forme et fonction de la définition en discours*, PhD Thesis, Université Toulouse II-Le Mirail.
- RIDDOCH M. Jane and HUMPHREYS Glyn W. (1987a): “Visual object processing in optic aphasia: A case of semantic access agnosia”, *Cognitive Neuropsychology*, 4, p. 131-185.
- RIDDOCH M. Jane and HUMPHREYS Glyn W. (1987b): “A Case Of Integrative Visual Agnosia”, *Brain*, 110, p. 1431-1462.
- RIEGEL Martin (1990): “La définition, acte du langage ordinaire: De la forme aux interprétations”, *La définition*, Eds CHAURAND J. and MAZIÈRE F., Paris, Librairie Larousse, p. 97-110.
- ROBINSON Richard (1950): *Definition*, Oxford, Clarendon Press.
- RUBIN Daniel L., SHAH Nigam H. and NOY Natalya F. (2007): “Biomedical ontologies: a functional perspective”, *Briefings in Bioinformatics*, 9, p. 75-90.
- RUSSELL Bertrand (1905): “On Denoting”, *Mind*, 14, p. 479-493.
- SAGER Juan (1990): *A practical course in terminology processing*, Amsterdam, Philadelphia, John Benjamins.
- SCHOBEL Daniel, SMITH Barry, LEWIS Suzanna E., KUSNIERCZYK Waclaw, LOMAX Jane, MUNGALL Chris, TAYLOR Chris F., ROCCA-SERRA Philippe and SANSONE Susanna-Assunta (2009): “Survey-Based Naming Conventions for Use in OBO Foundry Ontology Development”, *BMC Bioinformatics*, 10, p. 1-9.
- SCHULZ Stefan, BALKANYI Laszlo, CORNET Ronald and BODENREIDER Olivier (2013): “From Concept Representations to Ontologies: A Paradigm Shift in Health Informatics?”, *Healthcare Informatics Research*, 19, p. 235-242.
- SEPPÄLÄ Selja (2006): “Semi-Automatic Checking of Terminographic Definitions”, *International Workshop on Terminology design: quality criteria and evaluation methods (TermEval) – LREC 2006*, Genoa, Italy, p. 22-27.
- (2012): *Contraintes sur la sélection des informations dans les définitions terminographiques: vers des modèles relationnels génériques pertinents*, PhD Thesis, Département de traitement informatique multilingue (TIM), Faculté de traduction et d'interprétation, Université de Genève.
- (2015): “An Ontological Framework for Modeling the Contents of Definitions”, *Terminology*, 21, p. 23-50.
- SEPPÄLÄ Selja, SCHREIBER Yonatan and RUTTENBERG Alan (2014): “Textual and logical definitions in ontologies”, *Proceedings of the First International Workshop on Drug Interaction Knowledge Management (DIKR 2014)*, *The Second International Workshop on Definitions in Ontologies (IWOOD 2014)*, and

- The Starting an OBI-based Biobank Ontology Workshop (OBIB 2014)*, Eds BOYCE R. D., BROCHHAUSEN M., EMPEY P. E., HAENDEL M., HOGAN W. R., MALONE D. C., RAY P., RUTTENBERG A., SEPPÄLÄ S., STOECKER C. J. and ZHENG J. Houston, TX, USA, CEUR Workshop Proceedings (CEUR-WS.org, vol. 1309), p. 35-41.
- SHALLICE Tim (1988): *From neuropsychology to mental structure*, Cambridge University Press, Cambridge ; New York.
- SMITH Barry (2013): “Introduction to the Logic of Definitions”, *Proceedings of the International Workshop on Definitions in Ontologies (DO 2013) in Proceedings of the 4th International Conference on Biomedical Ontology Workshops (ICBO 2013)*, Vol-1061, Eds SEPPÄLÄ S. and RUTTENBERG A., Montreal, Canada, CEUR Workshop Proceedings (CEUR-WS.org), vol. 1515.
- SMITH Barry and CEUSTERS Werner (2015): “Aboutness: Towards Foundations for the Information Artifact Ontology”, *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO2015)*, Lisbon, Portugal (CEUR Vol-1515), 2015, p. 1-5.
- SMITH Barry, CEUSTERS Werner, KLAGGES Bert, KÖHLER Jacob, KUMAR Anand, LOMAX Jane, MUNGALL Chris, NEUHAUS Fabian, RECTOR Alan and ROSSE Cornelius (2005): “Relations in biomedical ontologies”, *Genome Biology* 6, R46.
- SMITH Barry, KUSNIERCZYK Waclaw, SCHOBER Daniel and CEUSTERS Werner (2006): “Towards a reference terminology for ontology research and development in the biomedical domain”, *Proceedings of KR-MED*, p. 57-66.
- STEVENS Robert, MALONE James, WILLIAMS Sandra, POWER Richard and THIRD Alan (2011): “Automating generation of textual class definitions from OWL to English”, *Journal of Biomedical Semantics*, 2, p. S5.
- VÉZINA Robert, DARRAS Xavier, BÉDARD Jean and LAPOINTE-GIGUÈRE Micheline (2009): *La rédaction de définitions terminologiques*, Montréal, Office québécois de la langue française.
- WARRINGTON Elizabeth K. (1985): “Agnosia: The Impairment of Object Recognition”, *Handbook of Clinical Neurology*, 45, p. 333-349.
- WEINREICH Uriel (1970): “La définition lexicographique dans la sémantique descriptive”, *Langages*, 5, p. 69-86.

DEFINITIONS IN ONTOLOGIES

Abstract: Ontologies standardly contain definitions, which can be of two kinds: *textual*, to help human users understand and correctly use the ontology terms; and *logical*, to allow automated systems to check the consistency of the ontology, to enhance querying, and to integrate and compare data annotated using the ontology. Textual and logical definitions have overlapping and complementary functions. Both share the functions of fixing, clarifying, and conveying the intended meaning of terms in the ontology. Both fulfil a cognitive function by enhancing the lexical competence of users when working with the ontology terms whose meanings are specified.

In this communication, we examine the different kinds of things to which the term ‘definition’ refers, and review their linguistic and cognitive functions. We focus on the function of definitions as a means for enhancing the semantic inferential competences of human users and automated systems. We also emphasize functions more specific to definitions in ontologies. Finally, through comparisons of textual and logical definitions, we show how analyzing these correspondences may be useful for improving the quality of definitions in ontologies.

Keywords: ontology, definitions, textual definitions, logical definitions, functions of definitions, definition checking

LES DÉFINITIONS DANS LES ONTOLOGIES

Résumé: Les ontologies comportent des définitions qui, typiquement, y sont de deux sortes : textuelles et logiques. Les *définitions textuelles* aident l'utilisateur à comprendre et à utiliser correctement les termes de l'ontologie ; les *définitions logiques* permettent aux systèmes informatiques de vérifier la consistance logique des ontologies, d'enrichir les requêtes, et d'intégrer et comparer des données annotées à l'aide d'ontologies. Les définitions textuelles et logiques ont des fonctions partagées et complémentaires. Toutes deux partagent la fonction de fixer, clarifier et communiquer le sens voulu des termes de l'ontologie. Toutes deux se complètent dans leur fonction cognitive d'ajuster et, ainsi, améliorer les compétences lexicales des utilisateurs lorsqu'ils emploient les termes définis dans l'ontologie.

Dans cet article, nous examinons les différents types de choses auxquelles le terme ‘définition’ réfère et passons en revue les fonctions linguistiques et cognitives des définitions. Nous mettons l'accent sur leur fonction d'ajustement des compétences inférentielles sémantiques des utilisateurs et des systèmes. Nous examinons également les fonctions propres aux définitions dans les ontologies. À travers des comparaisons de définitions textuelles et logiques nous montrons, pour conclure, comment une analyse de leurs correspondances peut s'avérer utile pour améliorer la qualité des ontologies.

Mots-clés: ontologies, définitions, définitions textuelles, définitions logiques, fonctions des définitions, contrôle-qualité de définitions