# Foundations of Argumentation for Argument Mining

Laura Alonso Alemany, Patrick Saint-Dizier

University of Cordoba – IRIT - CNRS,
Cordoba, Argentina – Toulouse, France.
lauraalonsoalemany@gmail.com – stdizier@irit.fr

July 5, 2017

# Course Outline

**(1) Conceptual and Linguistic foundations of argumentation**, Argumentation models in language and communication,

**(2) Formal and Computational models for argumentation**, Computational representations, argumentation schemes, argumentation and domain knowledge,

**(3) Corpus compilation and annotation**, Methodology for (argumentative) corpus annotation, An overview of annotated corpora,

**(4) Argument analysis, argumentation construction**, Functional architecture, applications,

**(5) Discussion**: concepts, implementations, applications and difficulties, interesting areas of work and open questions.

# A few Major General References

- J. Moeschler, argumentation and conversation, Hatier, translation. 1985,
- Van Eemeren et ali., Crucial concepts in argumentation, Amsterdam univ. Press, 2001,
- Van Eemeren et ali. Argumentation, Routledge, 2010
- C. Plantin, L'argumentation, Seuil, 1996
- D. Walton et ali. Argumentation schemes, CUP, 2008,
- P. Besnard, A. Hunter, Elements of argumentation, MIT Press, 2008,
- J. Searle, Speech acts, an essay in the philosophy of language, CUP, 1969.

# A- DEFINITIONS, AIMS
## What is argumentation ?

- Argumentation is essentially aimed at (1) convincing someone of a certain point of view or (2) coming to a reasonable agreement between two or more parties about a disagreement (controversial issue or standpoint).

- Difference of opinion: parties express doubts about the other party's standpoint, oppositions may bear on a unique or on multiple points,

- argumentation can be dialogical or based on written elements,

- argumentation is both normative and descriptive: identify differences of opinion and how they can be resolved.

- argumentation is expressed via language and possibly via non-verbal forms. Argumentation makes a heavy use of rhetoric.

# Application areas: some prototypical ones

- opinion analysis: reaching the why, and value systems,

- debate analysis (oral, written),

- business intelligence (weak signals with arguments),

- decision making (paired with a decision theory),

- population trends / values analysis on the long term,

- juridical defenses, pleads, etc. mediation, deliberations,

- detection of incoherences in sets of arguments (juridical, technical, etc.).

# Simple Illustrations: standpoints or controversial issues

*(1) Ebola vaccination is necessary,*
*(2) Women's conditions have improved in India,*

*(3) The development of nuclear plants is necessary,*

*(4) Organic agriculture is the future,*

*(5) Product P quality and features F,*

*(6) Political decision D and consequences C*.

# Illustration: an argument and its environment

Argument kernel + discourse structures around (discourse modifiers):

*<argument>*
*<concession> Even if the vaccine seems 100% efficient and without any side effects on the tested population, </concession>*

**<main arg> it is necessary to wait for more conclusive data before making large vaccination campaigns </main arg>**

*<elaboration> The national authority of Guinea has approved the continuation of the tests on targeted populations.</elaboration>*
*</argument>.*

# arguments for or against issue (1)

**Supports:** *vaccine protection is very good; Ebola is a dangerous disease; high contamination risks; vaccine has limited side-effects; no medical alternative to vaccine, etc.*

**Attacks:** *limited number of cases and deaths compared to other diseases; limited risks of contamination, ignorance of contamination forms; competent staff and P4 lab difficult to develop; vaccine toxicity and high side-effects,*

**weaker forms: Concessions or Contrasts:** *some side-effects; high production and development costs; vaccine not yet available; ethical and freedom problems.*

# arguments for or against issue (2)

**Supports:** *increased percentage of literacy among women; women are allowed to enter into new professional fields; at the upper primary level, the enrollment increased from 0.5 million girls to 22.7 million girls.*

**Attacks:** *practices of female infanticide, poor health conditions and lack of education still persisting; home is women's real domain; they are suffering the violence afflicted on them by their own family members; malnutrition is still endemic.*

# arguments for or against issue (3)

**Supports:** *allows energy independence; creates high technology jobs; risks are over-estimated; wastes are well-managed and controlled by AIEA; preserves the other natural resources,*

**Attacks:** *there are alternative solutions with less pollution; alternatives create more jobs than nuclear; there are risks of military uses: more dangerous than claimed; nuclear plants have high maintenance costs, etc.*

**Concessions or Contrasts:** *nuclear plants use dangerous products, but we know how to manage them; difficult to manage nuclear plants, but we have competent persons,*

Arguments require knowledge about energy, energy management, pollution, health, and various related dangers.

# Argumentation: a network

- arguments may attack or support each other: *nuclear plants use dangerous products / we know how to manage them*

- arguments may attack the inference that connect two arguments: *Ebola is dangerous with high contamination, therefore vaccination is necessary / vaccination does not stop dissemination.*

# The Origins of Argumentation

- Origin seems to parallel the development of geometry.

- Origin attributed to Tisias and Corax, probably Aristotle (384-322 BC) and sophists (5th century BC): critiques of their society.

Some features of the Greek argumentation tradition:

- Antiphonia: game: given a discourse, produce a counter discourse.

- Notion of probable (and associated forms of paradoxes), from which emerged notions such as prototypes and types.

- Dialectic interactions, critique of natural language as a means to establish forms of scientific truth.

# Argumentation and Rhetoric: the argumentative process (1)

Inverse perspective: argumentation is considered to be the foundational component of rhetoric.

discourse = set of language acts, that follow a plan. A discourse always has a goal.

- Main types of argumentative discourse: deliberative, judiciary, epidictic, exhortation, epistolary, advertising, propaganda.

- global traditional structure of the argumentative process: invention, disposition/planning, elocution, memorization, action.

# Argumentation and Rhetoric: the product (2)

Argumentative discourse structure from a rhetoric perspective:

- introduction (exordium)

- narration of facts, from a certain standpoint

- argumentation (defense), with its codes and processes,

- refutation by opponents,

- conclusion, summary of main points.

$\Rightarrow$ Scientific argumentation is based on forms of logic, besides rhetoric. BUT argumentation is weaker than a proof! (e.g. *I need proofs, Sir, not arguments!*).

# Contemporary trends

Argumentation uses foundational works by: J.L. Austin, J.R. Searle (language acts), Grice (Logic and conversation), etc.

Five main theoretical research directions (partly from C. Perelman, 1958):

- Pragma-dialectics (Van Eemeren et ali.): argumentation as a type of dialogue following strong norms,

- Argumentation and Conversation (J. Moeschler, E. Roulet): analysis of verbal interactions,

- Pragmatic and Linguistics (Anscombre and Ducrot)

- The communicational act (Habermas): ethics of argumentation,

- Logical Pragmatics (Grize): development of natural logics, cognition.

Applications: Plantin, Toulmin, Walton, Willard, Van Eemeren, etc.

# Theoretical intersections

- thought / language: argumentation is a mental process associated with a linguistic activity,

- language / discourse: Most / all statements have an argumentative effect on the listener, they affect his thought system, however only some discourses are argumentative due to their internal organization,

- monologue / dialogue: debate vs. essay: structures, goals, and organization largely differ,

- argumentation follows / does not follow norms. A norm allows to decide wether an argumentation is sound. Types of norms: efficiency, accuracy, truth.

- consensus / dissensus: construct a consensus around a standpoint, vs. identify divergences between opponents: powerful means to develop critical thinking.

# B- The BASIC SCHEMES
## Standpoints and the Argument-conclusion pair

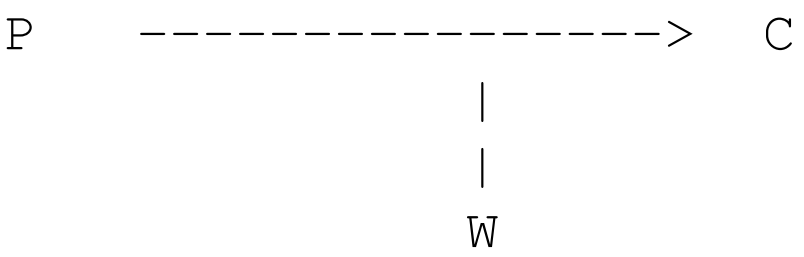
- Standpoints (E2 below, also called conclusions) are basically rational statements, they can be: the expression of a fact or an event, a prediction, a judgment, an advice or a warning.

- Whatever form a standpoint has, it is always possible to express some form of doubt about it.

Given two utterances E1 and E2, the relation:
Argument (E1) - Conclusion (E2)
can be realized explicitly by clues such as:
*E1 motivates, defends, induces to believe, implies, therefore, causes, explains, proves, E2* etc.
or, conversely:
*E2 given, because... E1*.

# The elementary argumentative schema, Toulmin's model (1)

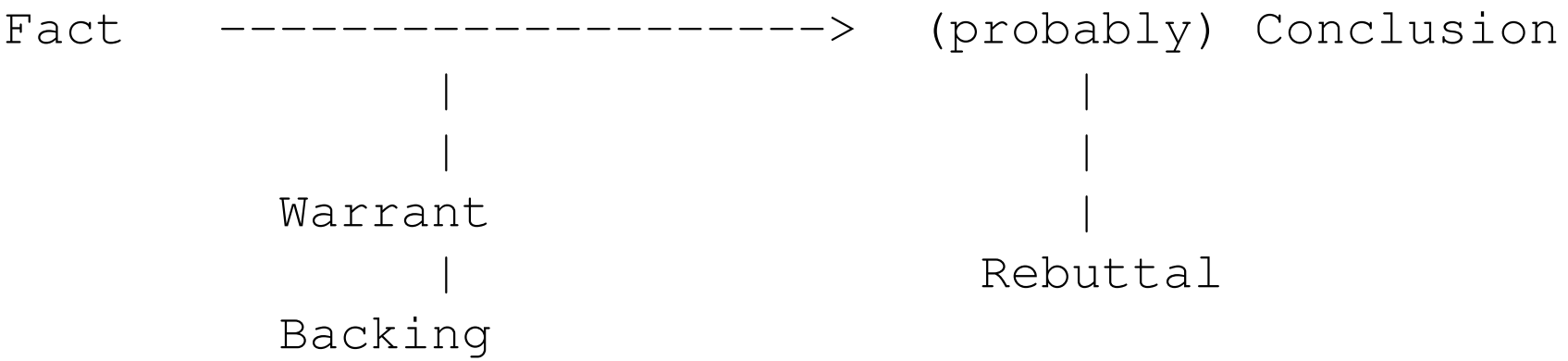
The argumentative cell:
P $\Rightarrow$ C. (Premise / Conclusion; E1 / E2)
It freezes this morning $\Rightarrow$ flowers will suffer.
With the 'passage' rule (Warrant):




e.g. W = frost deteriorates flowers.

# The elementary argumentative schema, Toulmin's model (2)

The typical argumentative cell:




Rebuttal: expresses forms of restrictions of the warrant:
*Fact: It freezes this morning*
*Conclusion: flowers will suffer*
*Warrant: plants in general do not like frost*
*Backing: physical law in botanic*
*Rebuttal: unless they got early sun or were well protected....*

# The elementary argumentative schema, Toulmin's model (3)

Some terminology:

Claim / Conclusion: the conclusion or argument being made

Grounds / Facts / Evidence: the data and facts offered to support the claim

Warrant: connects the ground to the claim

Backing: supports / explains the warrant

Qualifiers: statement about the strength of the claim

Rebuttal: exceptions to the claim.

Argumentation: starts from a fact (the argument) to reach another statement (the conclusion) that is less certain.

# Argumentation vs. Demonstration (1)

- If an argumentation is based on scientific (logical) norms N: it is a demonstration.

- Discussions on the validity of the norms N: basis of the notion of paralogism.

- paralogism (fallacy): argumentation that does not meet the rules of syllogisms, but has a form that resembles a valid argumentation: *The French are often on strike, Some professors are French, therefore some profs are on strike.*

- some typical forms of fallacies:
* of ambiguity: A is B, C is D, therefore A is D, where B and C are the same term but with a different facet, e.g. book.
* of deduction: several forms, e.g. *A genius is never understood, no one understands me, therefore I'm a genius.*

# Argumentation vs. Demonstration (2): Paralogisms / Fallacies

- Misuse of a formal deduction rule or technique, e.g. in predicate logic, relational algebra, etc.
- misuse of induction rules,
- incorrect form of analogy, when reasoning by analogy,
- incorrect conclusion from statistical analysis or probability computation,
- false observations,
- incorrect beliefs (?).

Main question: how argumentation must be driven, at what level of scientific analysis ?

# C- STRUCTURE of ARGUMENTATION
## Typologies of argumentation (1)

- Perelman et ali. (1958): quasi-logical argumentations, arg. based on the structure of reality (relations between objects, causality, etc.), inferences structure reality (analogy, exemple),

- Toulmin et ali. (1984): 9 types of arguments, based on: analogy, generalization, cause, authority, opposites, degree, etc. kind of argumentation schemes.

- Van Eemeren et ali. (1992): 10 rules of critical discussion: 'argumentation contract'.

# Dimensions for argumentation

- Argumentation uses objects or facts and relations between objects or facts (causality, analogy, functions, etc.), argumentation is based on the structure of reality, as percieved by listeners and speakers,

- Argumentation undergoes language constraints and their possible effects / side-effects / distortions / ambiguities / implicitness,

- Argumentation is basically an interactive process (interactions, ad'hominem, authority, etc.),

- Argumentation is not deduction (resulting in many difficulties), it follows various types of argumentation schemes, it is not a monotonic process.

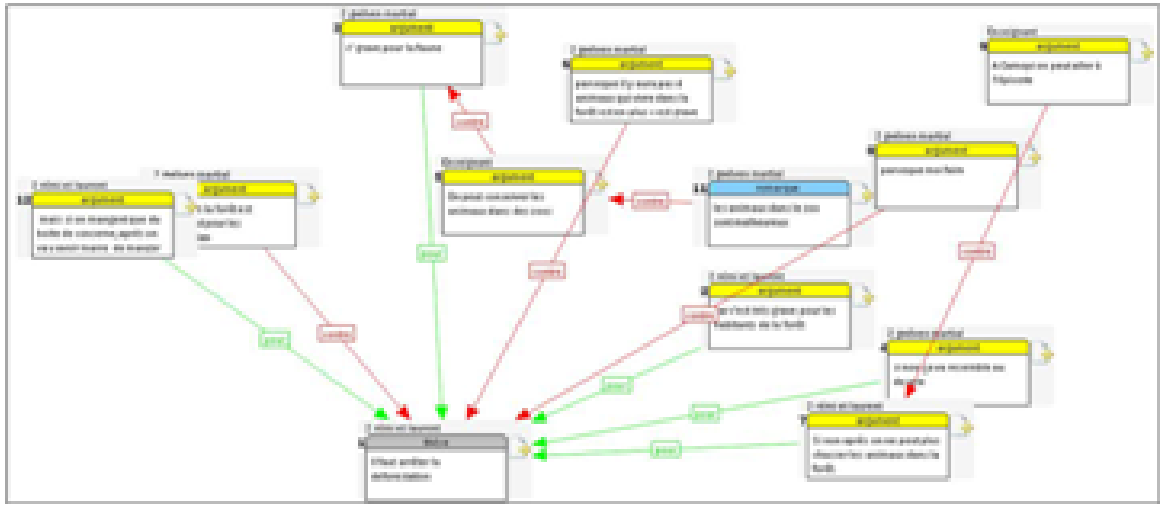
# Arguments and argumentation

- in a debate, a new disagreement may arise at any point and originate new doubt expressions,

- separate main versus subordinate difference of opinion,

- an argument is the expression of a doubt (attack) or of an approval (support) w.r.t. a statement, a part/facet of a statement, or another argument, example: *vaccination prevents bio-terrorism*.

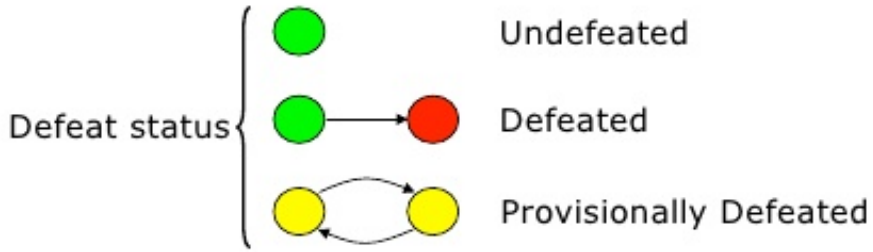
- an argumentation is a graph that, given a statement, represents the attack or support relations with the issue and between arguments:



## Argumentation semantics

- Specification of a method for argument evaluation, or of criteria to determine, given a set of arguments, their "defeat status"


*Argumentation Framework* *Defeat status*




25

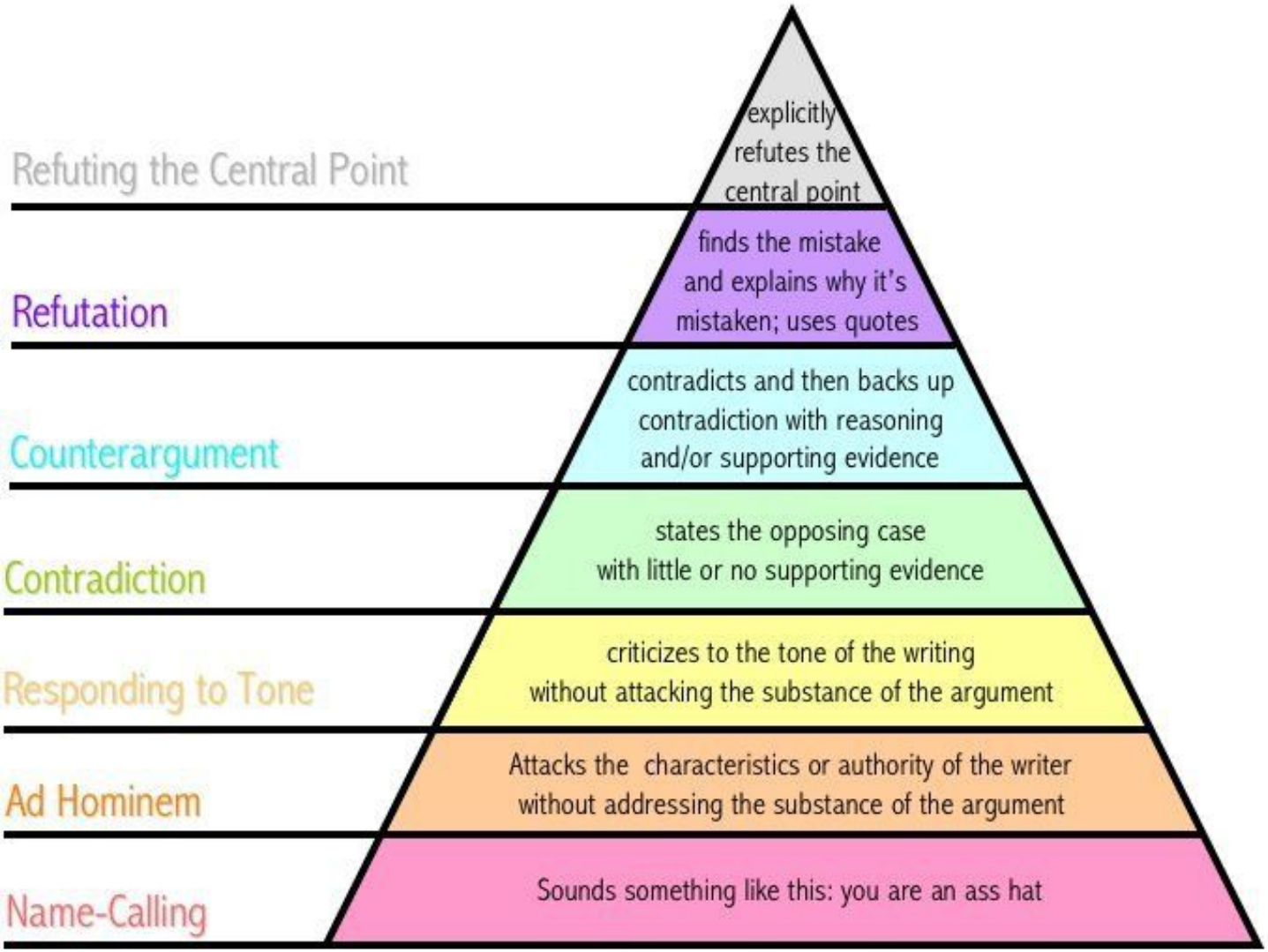
# Forms of disagreements, attacks: examples

Vaccine is too expensive for poor countries / when large volumes will be produced, it will be much cheaper and accessible to everyone.

Nuclear: there are risks of military uses / the AIEA controls wastes and production

Women's conditions: practices of female infanticide still persisting / but stronger and independent control has been implemented in each village via nurses ...

# Hierarchy for disagreement



# Global Organization of an argumentation

- via argumentation (discussion, debate, forums, etc.): includes a protagonist and an antagonist, who challenges the standpoint, and possibly vice-versa. Sometimes use of a moderator (debates, mediation).

- goal: reach an agreement, or convince of the well-foundedness of a standpoint or of its opposite.

- modern and general, 'ideal' model of an argumentation: **confrontation** (recognize difference of opinion), **opening stage** (decide to try to resolve the pb), **argumentation stage** (defense of standpoints via argumentation rules and conventions), **conclusion stage** (assess the differences or agree on a compromise).

-argumentation vs explanation, may co-exist. Explanation includes RST structures such as elaboration, clarification, illustration or cause. Justification is different. These can form secondary standpoints.

# Unexpressed standpoints and premises

- Communication rules impose to make explicit every element in the discussion. However, implicit and/or unexpressed elements (premises or standpoints) may arise and are part of the argumentation strategy.

- Antagonists may infer different unexpressed or implicit standpoints. *The vaccine adjuvant is toxic because P*: P may cover various reasons.

- All elements of an argumentation are crucial to evaluate its soundness.

- Communication rules principles: be clear, be sincere, be efficient, keep to the point. Still allow unexpressed elements but make them more ‘accessible’ to the other party.

- These rules should apply to any type of speech act.

# The complexity of argumentation

- Defense of a standpoint (supports) by means of one or more arguments. For a better evaluation, make explicit all premises,

- Similarly, expression of doubts (attacks) is often made via several arguments, which must be as explicit as possible.

- Multiple argumentation: use of several alternatives to defend or attack a standpoint, these do not depend on each other, their weight may be different.

- Coordinative argumentation: conjunction of arguments that form an homogeneous whole, they depend on each other in several ways, their structure is flat,

- Subordinative arguments: chain of arguments: arg. i+1 supports arg i. structured as layers. Some arguments in the chain become sub-standpoints related to a sub-argumentation. The weakest arg determines the strength of the whole.

- Argument acceptability is a matter of degree: more arguments entail a higher acceptability.

# Examples of the various types of argumentation structures

Multiple: C: You cannot get vaccined against Ebola in Monrovia; P: vaccine against Ebola is not available ; Monrovia is not accessible to visitors.

Coordinative: C: we had to go to a restaurant for dinner; P: the fridge was empty AND all stores were closed.

Subordinative: C: I cannot help you to repaint your house; C: I have no time ; C1: I have to prepare exams; C2 otherwise I will loose my financial support, etc.

These can be mixed.

# Linguistic clues in complex argumentation

How to detect from language the type of argumentation structure ? Not always explicit, language may help, knowledge (domain, common-sense) may greatly help.

In a well constructed argumentation, connectors and other clues may help, some typical clues:

- Multiple arg.: *apart from, not to mention, another reason, aside from,* etc.

- Coordinative arg.: *as well as, in addition, on top of that, especially, not only, more importantly,* etc.

- Subordinative arg.: *because, therefore, since, that is why,* etc.

- Some closing clues: *I conclude that, taking everything into consideration, all things considered, ...*

- Identifying the type of arg. is crucial to evaluate it.

# D1 - ARGUMENTATIVE DISCOURSE EVALUATION

- Objective: check for pragmatic and logical inconsistencies: each arg must be valid, make explicit unexpressed args.

- identify argument schemes that are used, linking premises to standpoints. Evaluate the validity of schemes in the current context (e.g. generalizations, arg from authority, etc.)

- Several types of incorrect argumentative discourse: each element of an argument must be acceptable, logically valid and the arg. scheme used must be appropriate, otherwise the argumentation is not fully acceptable.

- Soundness of argumentation depends on how schemes are used, in particular schemes based on: symptomatic relations, analogies and causality,

# Classes of argument schemes

- schemes based on symptomatic relations: standpoint defended in arg by citing a sign, a symptom or a specific mark typical of the arg. $\Rightarrow$ typical feature, property, quality, etc.

- schemes based on analogies: something in the standpoint is also cited or generalized in the argument(s): the standpoint should be accepted from this analogy. an element X in the standpoint and another element Y in the arg share many characteristics.

- schemes based on causality: a standpoint is defended by a causal connection with an argument, such that the standpoint should be accepted on the basis of that connection.

# Argument schemes (1)

More developed in D2 and in second part. A few typical examples:

- Argument from Analogy, developed below,

- Argument from Cause to Effect: *generally if A occurs then B occurs. How strong is the causal link ?*

- Argument from Correlation to Causes,

- Argument from Established Rule: *for all X if A is the rule for X, then (except for specific cases) A must do X. Is that possible that there are several rules $A_i$: then doubts on the most appropriate one.*

- Argument from Evidence to a Hypothesis: *if A (hypothesis) is true then B (a fact, an event) will be observed to be true. Could be some other reason for B to be true not related to A ?*

# Argument schemes (2)

- Argument from Example: *A has properties F and G, A is typical of things/events that have both F and G. Does this supports a general claim ? is the example typical ? how strong is the generalization ?*

- Argument from Popular Practice: *a large majority does A or act as A is the right thing to do, same for popular attitude: most people think A: A is probably right.*

- Argument from Expert Opinion

- Argument from Precedent.

# Fallacies

Fallacies are violations of the rules of argumentation and critical discussion principles. There are many types, among which, most notably:

- making pressure or attacking the opponent,

- shifting the burden of proof by debating other topics,

- defining a kind of straw man

- using fake arguments

- using rhetorical effects to confuse the opponent,

- denying an unexpressed premise.

# Dialogically and Conceptually avoiding fallacies (1)

Rules to resolve a divergence:

- do not prevent the other partie(s) to give standpoints or express doubts: do not: put pressure or discredit on the others, do not attack credibility, integrity, personal interest, etc. ad'hominem, tu quoque,

- when a standpoint is uttered, it must be defended by its utterer if asked to do so: must not shift the burden of proof, organize sequences of standpoints If there are several, do not forget any, start by the most straightforward.

# Dialogically and Conceptually avoiding fallacies (2)

- an attack or a support must always be clearly related to a standpoint that has been uttered: do not shift in the topics, do not use fake doxa, do not misrepresent or simulate a fictitious standpoint uttered by the other party (straw man): it is easier to attack than the original one.

- defend only your own standpoint: develop relevant arguments, no abuse of authority, otherwise: irrelevant argumentation, which may rather play on the pathos (pathetic fallacy)

- do not falsify or deny an unexpressed premise that is not really ambiguous or vague: implicit elements are frequent in communication, there are many implicit premises, but they must be well-known to everyone or easy to infer.

# Dialogically and Conceptually avoiding fallacies (3)

- do not violate a starting point: do not present a premise which is not the accepted starting point, or do not deny an accepted starting point: partie smust share a minimum of common behavior rules, beliefs, norms and value hierarchies,

- apply an appropriate argument scheme otherwise a standpoint or an argument may not be conclusively defended, do not use inappropriate schemes (populist fallacy, confuse facts with judgements, hasty generalization, false analogy, incorrect cause-effect relation, etc.),

- the reasoning model used in the argumentation must be logically valid (e.g. do not confuse necessary and sufficient conditions)

# Dialogically and Conceptually avoiding fallacies (4)

- A defense that fails results in the retraction of the corresponding standpoint,

- a defense that succeeds must result in the retraction of the antagonist doubts, but frequently not so boolean: facets may succeed.

- when both fail: elaborate a kind of middle course.

- principle of cooperation: only use clear, as unambiguous as possible formulations, in order to avoid confusions and misinterpretations. Avoid: implicit elements, indefinite references, unfamiliar terms and vague or underspecfied concepts or constructions.

# D2 - SPECIFIC FORMS of ARGUMENTATION argumentation and analogy (1)

Scheme: P is being debated. P' is is true. P and P' share a lot of similarities. Therefore P is true.

Rather vague, with little explicative power.

Various forms to express analogy: P, similarly to P', P and P' are of the same type, etc.

*P: In an introduction, examples must be simplified*
*Opposition: simplified examples are useless*
*analogy with P: P': examples in medical introductory courses are recognized to be helpful*
*Refutation by analogy: in this course, simplified examples are useful.*

# argumentation and analogy (2)

Refutation of the analogy rule: everything is (vaguely) analogous to everything.

ad'hominem character of analogy: opponent accepts analogy, but goes further by elaborating a point in the analogy that contradicts the conclusion.

*P: should methadone be refunded by social services ?*
*- yes, because treatments against alcohol are also refunded*
*- ad'hominem refutation: no, because methadone is not a treatment but a drug.*

Types of analogies in argumentation: in the juridical domain: reference to previous and similar cases (jurisprudence), in political discourses: parangons (major historical situations), moral cases.

*my sister offered me my favorite cartoon, it is normal that I help her to prepare her maths exams.*

# Argumentation on the nature of entities (1)

When a definition is clear and sound, it can be used as a scheme for an argumentation based on definitions: articulation point between objects and language in argumentation.

schema: a notion N is defined by means of features.

Then, for an entity X : does it belong to N ?.

The properties of X are compared / intersected with those of N, via a possible debate on the nature of things and how their are described in language.

On that basis, X is or is not an element of N.

$\Rightarrow$ Argumentation by essence.

# Argumentation on the nature of entities (2)

- Definition must be established a priori, independently of any case to debate.

- avoid ad'hoc definitions, based on precise entities X: bias to the argumentation scheme.

- argumentative definition: includes a position: *university research is disconnected from reality*. May lead to negative conclusions on this type of research!

# E- ARGUMENTATION and RHETORIC
## Rhetoric and communication

- demonstration: impersonal mode, based on facts, axioms and inference rules. Various types of logics and forms of demonstrations.

- dialectic: art of dialog, doesn't deal with truth, but with what is probable. However, use of rigorous reasoning to reach the best consensus on a topic. Use argumentation, negotiation, mediation, etc.

- sophism: area of duplicity and cheating, based on false assumptions which nevertheless seem plausible to an audience.

Structure of a classical discourse based on rhetoric: invention, arrangement, style, delivery (incl. vocals, mimics).

# The elocution steps in classical rhetoric

- main features: correctness (lexical choice, syntax w.r.t. genre: judicial, deliberative, epidictic, etc.), clarity (no ambiguity), pleasant to listen (metaphors, figures of speech),

- appropriate style: simple (logos, proof, etc.), middle (ethos: give a good impression to the audience), noble (pathos, for persuasion).

- figures of speech: of words (play with words for effects), of construction (ellipsis), of meaning (metaphors), of thought (relation between the orator, its topic, and the discourse: irony, emphasis).

# Elements of Pragmatics for argumentation analysis: Logos (1)

Logos: art of reasoning, of constructing demonstrations.

- Essential topics in language: argumentative value of negation, of interrogative forms (e.g. rhetorical questions); syntactic forms and focus, reported speech and citations, modals, etc.

- strategies of argumentation structure: sequencing arguments: role of implicit data and of connectors.

- lexical choice and its argumentative weight: e.g. word connotations (wild vs. un-controlled), induced irony, a priori judgments or polarity, etc.

- lexical re-interpretations, semantic variations: war vs. conflict of interest vs. men which fight each other (if we stop fighting: no war): radically changes perspectives.

# Logos (2)

- implicit aspects: leaves the opponent make his own inferences. Very strong argumentation mode: implicit = this is obvious, you agree with it. Analyze shared knowledge and beliefs to have a better view of argumentation.

- presuppositions: can often be reconstructed linguistically. More objective than implicit elements: has a persuasion effect. Reconstructed via e.g. Grice's implicature system, when presuppositions obey cooperation maxims. Useful to reconstruct the missing premises of an argumentation.

- connectors: bind 2 explicit statements or implicit ones with explicit ones: but, however, because, nevertheless, etc. they have their own pragmatic profile and implications.

# Ethos

crucial in dialogs, Ethos mainly deals with the characteristics of the orator: his behavior, his personality, how he speaks, etc. Elocution steps.

- a major parameter of persuasion (strength in context): consequence: develop emotion, empathy, seduction of the opponent or audience. Call to his imagination, desires, etc.

- the speaker must have a good analysis of his audience and of the psychological profile of his opponent to behave 'correctly' and to follow the audience expectations.

- example in advertising: nice web page to make you dream....

# Pathos (1)

How to persuade an audience, how to touch the opponent: complement to the Ethos.

- act on opponents: touch them, make them become angry, frighten them, etc. Not necessarily very honest (opinion analysis): convince (rational) + persuade (irrational). But emotions are also value judgments in a cognitive perspective.

- risks: persuasion may deeply affect the rationality (Logos) of argumentation. feeling of trickery.

- connection to ad'hominen and ad populum types of arguments, which deeply affect the Logos.

- Pathos must be paired with moral values to get all its role which is crucial. These values depend on the argumentation framework: negotiation, mediation, convince because of a danger (ad baculum).

- emotions can be argued and useful ! existence of the dichotomy logos / pathos ??

# Pathos (2)

- How to produce 'emotions' in an audience: not to be confused with the orator's emotions ?

- in relation of a set of beliefs (e.g. children = innocence), major social images,

- the same facts can produce very different emotions depending on how they are presented in an argumentation and how emotion is constructed.

- Similarly, rejection of emotions can be constructed, e.g. via typical terms, which re-orient the polarity of the arguments.

- Emotion lexicalized by specific linguistic and pragmatic marks, also style, emphasis, rhythm, repetitions, etc.

# F- LINKS between ARGUMENTATION and ARGUMENT MINING: difficulties

- Argument Mining is still in an early development stage: much more complex than e.g. factoid question answering. Somewhat close to WHY question-answering.

- Complexity: of language involved, need of various types of knowledge, depends on media, etc.

- Implicit elements, schemes, intentions, knowledge.

- Application areas: opinion analysis, juridical reasoning modeling, mediation, business intelligence, education, etc.

# Argument Mining: Methodological issues

- proceed gradually: given a standpoint, gradually identify arguments in isolation.

- Assume relative rationality of arguments, regularity of expression, etc.

- analyze their characteristics and annotate them. Stabilize features which are annotated and define guidelines which may be revised several times.

- at the moment: mining occurs essentially on surface linguistic (or typographic) considerations. Difficult to mine for argument articulations, attacks, support, etc.

- main approach: collect relevant corpora, observe, annotate, structure linguistic resources from observations (via learning, etc.). Consider specific types of arguments and / or specific types of applications and domains: characterize argument forms on this restricted areas.

# Argument Mining: Aims

- first, given a standpoint, automatically collect arguments for or against it,

- construct a synthesis when there are many arguments, difficulties: what facet arguments attack/support, define polarity is not straightforward, eliminate redundancies, target the right level of generality in the synthesis, etc.

- possibly derive parts of an argumentation graph,

- analyze user value system from arguments

- investigate which argument schemes are the most frequent.

# H- IMPACT OF CONTEXT and KNOWLEDGE in Arg MINING

The Relatedness Problem: Given a controversial issue:
$\Rightarrow$ besides linguistic aspects, domain knowledge + inferences are often required, example:

- Issue: *the situation of women has improved in India*,
- (a) *early morning, we now see long lines of happy young girls with school bags walking along the roads*
- (a) is a support of the issue,
- BUT it requires knowledge and inferencing to explicit the relationships between women's conditions and school bags

# The Relatedness Problem - 2

- Then: (b) *School buses must be provided so that schoolchildren do not reach the school totally exhausted after a long early morning walk.*
- (b) is an attack of (a) (*these young girls may not be so happy*)
- it is not an attack of the issue: the facet that is concerned in the relation between (b) and (a) does not concern women's conditions in particular.

$\Rightarrow$ knowledge and reasoning useful to establish **relateness and polarity:** the WHY, HOW and HOW MUCH.

# The Relatedness Problem - 3

Evidence from other corpora:

- (1) MM dialogues (with non-adjacent arguments) (Dundee):
  *Let's be straight, the situation in Kenya was a bloody dirty war.*
  is related to and involves an agreeing relation with:
  *I think there was horror at Mau Mau.*
  $\Rightarrow$ Needs knowledge about Mau Mau populations to induce agreeing.
- (2) DOT corpus (Dundee):
  *we will do everything we can to make sure the process is timely*
  is a rephrase of:
  *we are always looking to see what we fund next, what can we prioritize.*
  $\Rightarrow$ Needs knowledge about process organization to characterize a rephrase.

# Problem Analysis and Research questions

- How to construct a valid corpus? : define issues and manually search for arguments found in various types of texts
- How to tag arguments to explore and characterize the need of knowledge and reasoning ?
- How to categorize the knowledge involved ?
- How to pair NLP with KR for Argument Mining ?
- How to account for the diversity of arguments w.r.t. an issue ?

Research direction: The Qualia of the Generative Lexicon: a useful lexical and knowledge representation for argument mining?

# Corpus Construction: issues

(1) Ebola vaccination is necessary,
(2) Women's conditions have improved in India,
(3) The development of nuclear plants is necessary,
(4) Organic agriculture is the future.

The text fragments which are investigated are extracts from various sources where these issues are discussed, e.g.: newspaper articles and blogs from associations.

These are documents accessible to a large public, with no professional consideration, but arguments are very diverse, with various levels of abstraction in KR.

# Corpus Construction

| Issue | Corpus size (text extracts) | nb. of different arguments annotated | overlap rate |
|---|---|---|---|
| (1) | 16 texts, 8300 words | 50 | 4.7 |
| (2) | 10 texts, 4800 words | 27 | 4.5 |
| (3) | 7 texts, 5800 words | 31 | 3.3 |
| (4) | 23 texts, 6200 words | 22 | 3.8 |
| Total | 56 texts, 25100 words | 130 | 4.07 |

# Arguments and argument compounds

$\Rightarrow$ Arguments seldom come in isolation.

$\Rightarrow$ They are often articulated within a context that indicates e.g.: circumstances, restrictions, illustrations, concessions, comparisons, purposes, and various forms of elaborations.

$\Rightarrow$ We call such a form an **argument compound**, where the argument is the kernel: —-> allows for a larger diversity of arguments.

# Corpus Tagging

The following tags have been identified, but need to be further elaborated:

1. the **text span involved** that delimits the argument compound and its kernel,
2. the **polarity of the argument** w.r.t. the issue: support, attack, argumentative concession or contrast. Concessions and contrasts are both discourse structures and criteria to evaluate argument polarity.
3. the **conceptual relation(s) with the issue**,
4. the **knowledge involved**, to identify the argument: list of the main concepts used. These come preferably from a predefined domain ontology, or from the annotator intuitions, if none is available. This list may be quite informal, it nevertheless contributes to identify the nature of the knowledge involved.
5. the a priori **strength of the argument**,
6. the **discourse structures** associated with the argument kernel.

# Illustration for issue (1)

*<argument nb= 11, polarity= concession ,*
*relationToIssue= limited proofs of efficiency and safety of vaccination,*
*conceptsInvolved= efficiency measure, safety measures, test and evaluation methods,*
*strength= moderate>*

<concession> Even if the vaccine seems 100% efficient and without any side effects on the tested population,
< /concession>
<main arg> it is necessary to wait for more conclusive data before making large vaccination campaigns < /main arg>
<elaboration> The national authority of Guinea has approved the continuation of the tests on targeted populations.</elaboration> < /argument>.

# Main characteristics of the corpus

From our manual analysis, the following argument polarities are observed:

- attacks: 53 occurrences
- supports: 33,
- argumentative concessions: 21,
- argumentative contrasts: 19
- and undetermined: 4.

The corpus shows a tendency to argue against an issue:
- attacks and contrasts = 55%,
- supports and concessions = 41%.

# Evidence for knowledge for argument mining

Need of knowledge (estimate from annotator):
*nb of args that require knowledge / total nb of arguments .*

| Issue | need of knowledge nb of cases (rate) | total number of concepts involved (estimate) |
|---|---|---|
| (1) | 44 (88%) | 54 |
| (2) | 21 (77%) | 24 |
| (3) | 18 (58%) | 19 |
| (4) | 17 (77%) | 27 |
| Total | 100 (77%) | 124 |

# Main concepts used in argument kernels and their expression in language (issue 1)

**Supports:** *efficiency is very good, 100% protection; avoids or reduces dissemination of disease; limited side-effects,* etc.

**Attacks:** *limited number of cases and deaths compared to other diseases; limited risks of contamination, ignorance of contamination forms; toxicity and high side-effects,* etc.

**Concessions or Contrasts:** *some side-effects; high production and development costs; vaccine not yet available; ethical problems,* etc.

# Main concepts used in argument kernels and their expression in language (issue 1) - 2

*Vaccine* is the root of the system. Facets of this concept used in arguments:
(1) the parts of a vaccine: the *adjuvant* and the active principle;
(2) its super types: a vaccine is a kind of medicine; and
(3) the most central aspects are: its purposes, goals and consequences, and how it is created, tested and sold:

- the concepts of *side-effect* and *toxicity* are consequences of using a medicine;
- the concept of *contamination* is related to one of the purposes of a vaccine, namely to avoid *disease dissemination*;
- *production costs* are related to the creation and development of any product, including medicines and vaccines, etc.

# From Concepts to Knowledge Representation

The terms used in argument kernels concern: purposes, properties, parts, creation and development, etc. of the head terms of the issue or of derived concepts.

These are relatively well defined in the Generative Lexicon:

Vaccine(X):

$$\begin{bmatrix} \text{CONSTITUTIVE: } [\text{ACTIVE\_PRINCIPLE, ADJUVANT}], \\ \text{TELIC: } \begin{bmatrix} \text{MAIN: PROTECT\_FROM(X,Y,D), AVOID(X,DISSEMINATION(D))}, \\ \text{MEANS: INJECT(Z,X,Y)} \end{bmatrix}, \\ \text{FORMAL: } [\text{MEDICINE, ARTEFACT}], \\ \text{AGENTIVE : } [\text{DEVELOP(T,X), TEST(T,X), SELL(T,X)}] \end{bmatrix}$$

# Main concepts used in argument kernels and their expression in language (issue 2)

mainly involve comparisons with men's living conditions or refer to general principles of human welfare:

**Supports:** *increased percentage of literacy among women; women are allowed to enter into new professional fields; at the upper primary level, the enrollment increased from 0.5 million girls to 22.7 million girls.*

**Attacks:** *practices of female infanticide, poor health conditions and lack of education still persisting; home is women's real domain; they are suffering the violence afflicted on them by their own family members; malnutrition is still endemic.*

No concessions or contrasts have been observed: argument polarities are very clearcut, with a very positive or negative tonality, proper to highly controversial and overheated debates.

# Main concepts used in argument kernels and their expression in language (issue 2)- 2

More complex Qualia, a sample:
human(X):

$$\begin{bmatrix} \text{AGENTIVE} : \begin{bmatrix} \text{GET(X, FULL EDUCATION)}, \text{GROW(PARENTS, X)}, \\ \text{EDUCATE(PARENTS, X)}, \text{GET(X, JOB)}, \\ \text{GET(X, SECURITY)}, \text{GET(X, RIGHTS, HEALTH)} \end{bmatrix}, \\ \text{TELIC} : \begin{bmatrix} \text{SOCIO-ECONOMIC} : \text{ENTER(X, WORK CAREER)}, \\ \text{DEVELOP(X, INNOVATIONS PRODUCTS)}, \text{EARN(X, MONEY)}, \\ \text{CULTURAL, SCIENCE, EDUCATION} : \text{DEVELOP(X, CULTURE)} \\ \text{SOCIETY} : \text{CONTRIBUTE\_TO(X, SECURITY)} \\ \text{FAMILY} : \text{HAVE(X, FAMILY)}, \text{EDUCATE(X, CHILDREN)}, \\ \text{CHOOSE(X, HUSBAND/WIFE)} \\ \text{POLITICS} : \text{PARTICIPATE(X, PUBLIC LIFE)}, \\ \text{MAINTAIN(X, SAFETY OF POPULATION)} \end{bmatrix} \end{bmatrix}$$

# Main language realizations

The above arguments are expressed in various ways:
- **evaluative expressions:** *Vaccine development is very expensive*,
- **comparatives:** *number of sick people much smaller than for Malaria*.
- **facts related to properties of the main concept(s) of the issue:** *Vaccine is not yet available. There is no risk of dissemination*.
- **facts related to the consequences, purposes, uses or goals** of the issue: *vaccine prevents bio-terrorism*.

# Modeling the Diversity / the Generative expansion of arguments - an attempt

- Arguments attack or support specific facets of the concepts of the controversial issue (called **root concepts**).
  $\Box$ *(protect_from(X,Y, (infect(E1,ebola, Y)* $\Rightarrow$ *get_sick(E2,Y)*
  $\Rightarrow \Diamond$ *die(E3,Y)))* $\wedge$ *avoid(X,dissemination(ebola))*.
- Arguments may also attack or support concepts derived from these initial concepts (related to functions, parts, etc.).
- For example, they may attack properties or purposes of the adjuvant or of the protocols used to test the vaccine. Arguments must however remain functionally close to the root.

$\Rightarrow$ Develop a network of Concepts and their Qualias derived from those involved in the controversial issue, with a limited depth.

# Case study 1

**Utterance A1:** *The adjuvant of the Ebola vaccine is toxic*
Utterance A1 matches with the language pattern:
`[np, 'is', evaluative_expression]`

A1 negatively evaluates the adjuvant (lexical feature of the adjective 'toxic'), but it does not explicitly say anything about the vaccine.

Then, given: Adjuvant(Y,X1):

$$\begin{bmatrix} \text{FORMAL} : \left[\text{VACCINE, MEDICINE, CHEMICALS}\right], \\ \text{TELIC}: \left[\text{DILUTE(Y,X1), ALLOW(INJECT(X1,P))}\right] \end{bmatrix}$$

where Y is the adjuvant of X1, which is the active principle of the vaccine X. The role of an adjuvant is to dilute X1, and to allow to inject it to patients P.

The constitutive role of *vaccine(X)* says that the adjuvant is part of the vaccine. The Qualia of 'adjuvant' indicates that the active principle X1 is mixed by dilution with the adjuvant Y :

- Adjuvant Qualia: dilute(Y,X1) Y and X1 are mixed together to form a single entity, the vaccine X.
- *upwards inheritance of a property in a part-of relation*: if a (major) constitutive part K1 of an object K has a property P, then (probably) the entire object K has P:
  $has\_property(K1,P) \wedge part\_of(K1,K) \Rightarrow has\_property(K,P).$
- since Y and X1 are parts of X, then since Y is toxic for humans, it follows that X is also toxic for humans.

Therefore, A1 **attacks** the controversial issue.

This statement may also be interpreted as a **contrast** to the controversial issue: 'the vaccine is necessary BUT it is toxic'. (Winterstein 2012)

# **A2:** *Seven persons died during the Ebola vaccine tests*

- In the GL structure of vaccine(X), the 'test' activity is related to the agentive role.
- Axiomatization of the GL structure:
  *by definition, the agentive role is pre-telic: it occurs before the functions or the roles given in the telic role and their related properties are active:*

  $\forall\, P(E) \in \textit{agentive-role}, \forall\, Q(E1) \in \textit{telic-role}, E \leqslant E1 \wedge \neg(P \Rightarrow Q).$
- From that point of view, A2 is about tests, it does not say anything about the vaccine roles, functions and consequences once it has been fully tested and approved.
- Argument2 is **irrelevant or neutral** w.r.t. the controversial issue.

# What we learn from corpus analysis: Required knowledge and inference: a Synthesis

- **lexical knowledge:** semantic features for lexical items, in particular polarity, e.g. for verbs (avoid), intensifiers (for adverbs), scales, etc.,
- **domain knowledge:** encoded via the formalism of the Generative Lexicon, including event structures and causal chains, via a network of Qualia structures,
- **reasoning shemes:**
  1. inferences related to the semantics of the Qualia roles
  2. inferences related to lexical semantics structures,
  3. inferences related to general purpose or domain knowledge
  4. inferences dedicated to argumentation, that allow to compute relations and their strength between the controversial issue and the argument at stake. These are specific compositionality rules.

# K- NL SYNTHESIS of ARGUMENTS
# (1) Output of the mining system

*Even if the vaccine seems 100% efficient and without any side effects on the tested population, it is necessary to wait for more conclusive tests before making large vaccination campaigns. The national authority of Guinea has approved the continuation of the tests on targeted populations.*

<argument Id= 11, polarity= attack
conceptsInvolved= 'vaccine(X)/agentive/ test(T,X)'
strength= moderate >
<concession> *Even if the vaccine seems 100% efficient and without any side effects on the tested population,* </concession>
<main arg> *it is necessary to wait for more conclusive tests before making large vaccination campaigns.* </main arg>
<elaboration> *The national authority of Guinea has approved the continuation of the tests on targeted populations.* </elaboration>
</argument>.

# (2) Clusters of arguments

(1) The concepts in the network of Qualias are organized hierarchically and by role.

(2) Each of these concepts defines a cluster that contains related mined arguments.

(3) This process goes on with the lower levels of the network of Qualias.

(4) When an argument involves several concepts, it is included into the concept cluster that is the highest in the network of concepts.

(5) For each concept, related arguments are structured by polarity.

(6) Each argument that has been mined can also be an issue which can be attacked or supported.

# Cluster construction: Illustration for Issue (1)

**Level 1**: e.g.:
Cluster 1: Adjuvant: attack: *adjuvant is toxic* (3 occurrences) ...
Cluster 2: Dissemination: support: *reduces dissemination* (5)..
Cluster 3: Get-sick: concessive support: *limited number of cases and deaths compared to other diseases* (2), ...

**Level 2:**, e.g.:
Cluster 4: Production costs: attack: *high production and development costs (6)* ...
Cluster 5: Availability: concessive attack: *vaccine not yet available (4)*, etc.

$\Rightarrow$ Clusters are fine, but an overview is needed: Second step of the generation process.

# (3) A Higher Level Synthesis

Synthesis of arguments found from the network concepts + links to the set of clusters:

- from higher level / positive then negative polarities (frequency and link).

*Vaccine protection is good (3), bad (5).*
*Vaccine avoids (5), does not avoid (3) dissemination.*
*Vaccine is difficult (3) to develop.*
*Vaccine is (4) expensive.*
*Vaccine is not (1) available.*
*Ebola is (5) a dangerous disease.*
*Humans may die (1) from Ebola.*
*Tests of the vaccine show no (2), high (4) side-effects.*
*Other arguments (4).*

# Overview of the generation system (1)

synthesis produced via abstract linguistic patterns :

(1) `[HeadConcept, Be/Predicate, Evaluative, AttributeLexicalization].`
or:
(2) `[HeadConcept, Be/Predicate, AttributeLexicalization. Evaluative].`

`HeadConcept` is the lexicalization of the rightmost (or leaf) concept in the attribute 'conceptsInvolved'.

# Overview of the generation system (2)

`Be/Predicate` : 'be' (*is, are*) or *prevent, evaluate, allow, avoid*;

Main patterns:
Supports:
`[is/are/Verb, (Stats)],`

Attacks:
`[is not/ are not/do not Verb/ does not Verb, (Stats)]`

Supports and attacks:
`[is/are/Verb, (Stats1), is not/ are not/do not Verb/ does not Verb, (Stats2)]`

# Overview of the generation system (3)

`Evaluative`: scalar adjective, possibly modified by a negation depending on its polarity.

Adjective lexical choice :
(1) by default the values good / bad for products and attitudes and easy/difficult for processes.

(2) by an adjective found in one of the arguments of the cluster. The adjective must be prototypical or organized in non-branching proportional series (Cruse 1986).

toxicity : *poisonous - dangerous - neutral - recommended - beneficial*. prototypical: dangerous / beneficial, neutral: neutral.

# The current generation system

- This synthesis generation system is quite simple at the moment,

- lexical entries, related to the concepts in the Qualias : between 50 and 100 lexical entries depending on the issue,

- 22 patterns allow to produce the constructions presented above.

- Patterns are stable for the type of issues we consider, which are simple, and concrete,

- investigate: arguments may attack or support other arguments instead of the issue.

# Features of an Evaluation

- the overall linguistic adequacy of the generation system,
- the types of domains and controversial issues for which this system is adequate,
- the load in linguistic and conceptual resource
- the adequacy of the conceptual model, here the Qualia structure.
- the adequacy of this synthesis for professionals,
- user profile: there is no unique form of synthesis: several forms of synthesis could be foreseen that would depend on the reader's interests and profile.