Guidelines for Annotating Argumentation Structures in Persuasive Essays

Christian Stab and Iryna Gurevych

Ubiquitous Knowledge Processing Lab (UKP Lab)
Computer Science Department,
Technische Universität Darmstadt
Hochschulstrasse 10, D-64289 Darmstadt, Germany
www.ukp.tu-darmstadt.de

May 4, 2015
Version 1.2

Table of Contents

Chapter 1

Introduction

Argumentation Mining is an interdisciplinary research area that incorporates philosophy, psychology linguistics and computer science for establishing argumentation models and automated methods for identifying arguments in written texts. These tools will not only provide novel possibilities for educational applications like intelligent writing assistance, information retrieval platforms or automated assessment tools but will also open new opportunities for improving current legal information retrieval applications or policy modeling platforms. However, a major prerequisite for developing novel Natural Language Processing (NLP) methods that are able to identify argument components and argumentative relations in written texts is the availability of annotated corpora. Due to this requirement and the complex structure of argumentative discourse, "*the automatic detection of arguments has been left nearly unstudied*" till 2008 (Reed et al., 2008).

The goal of this study is to create a language resource for argumentation mining by manually annotating the structure of arguments in persuasive essays. Since the annotation an assessment of arguments is a complex task, this guideline first provides a brief introduction to argumentation theory including the definitions of argument components, argumentation structures and argumentative relations before describing the steps of the annotation in detail.

1.1 Arguments in a Nutshell

An argument consists of several statements. In its simplest form, it includes one claim that is supported by at least one premise (Peldszus and Stede, 2013; Britt and Larson, 2003; Toulmin, 1958). The *claim* (or also called conclusion Mochales-Palau and Moens (2009)) represents a controversial statement which the author tries to persuade the reader of. It is usually a proposition or assumption and should not be accepted by the reader without additional support. This characteristic distinguishes arguments from explanations where the conclusion is a true statement that is not arguable (e.g. an event that happened in the past). The second component of an argument, the *premise* (or sometimes called support (Besnard and Hunter, 2008) or reason (Britt and Larson, 2003)), underpins the plausibility of the claim. It is usually added by the proponent (writer) for persuading the reader of the claim. Considering the simplest form of an argument, a premise can be seen as a justification

for the claim, whereas more complex argumentation structures can also include premises that aim at refuting a claim. These more complex structures are basically graphs that connect premises and claims by means of different relations. Figure 1.1 illustrates the simplest form of an argument consisting of only one claim that is supported by one premise.

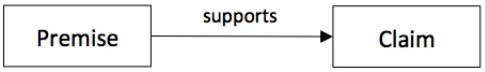

Figure 1.1: Simple form of an argument

Even such a simple form of an argument can be expressed in many different ways in written text. Some example patterns that can be found in written argumentation are the following:

<*claim*> because <*premise*>.
Since <*premise*> it is feasible that <*claim*>.
In view of the fact that <*premise*> it follows that <*claim*>.
<*premise*>. Therefore, <*claim*>.

However, there are many ways to express arguments in written texts and frequently the cue phrases (e.g. discourse markers like "*therefore*" or "*because*") are not present or misleadingly used in real texts. For instance, the following argument includes exactly one claim-premises-pair without any indicator:

> "*By wearing school uniforms, pupils are not able to develop their own style of fashion. Wearing school uniforms will have negative influence on the development of their characters.*"

In this example the second sentence is the claim which is supported by the first sentence but there is no indicator present which signals the argumentative relation between the two statements/sentences. The structures we will discover in persuasive essays are usually more complex and consist of several premises that either support or attack a certain claim. For instance, lets consider the following example:

> "*Although wearing school uniforms might foster the team spirit, it restricts the right of self-determination. Therefore, we should not force pupils to wear school uniforms.*"

This example includes three argument components (two in the first sentence and another one in the second sentence). The second sentence includes the claim "*we should not force pupils to wear school uniforms*" which is supported by the second premise in the first sentence "*it restricts the right of self-determination*". The first statement "*wearing school uniforms might foster the team spirit*" is a counter reason which attacks the claim. It is added by the author to make the argument stronger against any potential contra argument by an opponent. In this study we will annotate counter reasons as another premise which is connected to its target statement

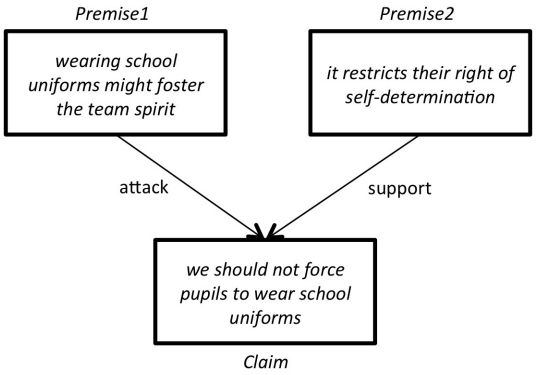

Figure 1.2: Example structure of an argument

(in this example the claim) with an attack relation. The argumentation structure includes therefore three statements and two argumentative relation. The structure of this argument is illustrated in figure 1.2: In the next section we will introduce the type of documents in which we will annotate argumentation structures and also describe some structural properties of these documents which might facilitate the annotation process.

1.2 Persuasive Essays

This annotation study is conducted on a set of persuasive essays. These essays usually exhibit a certain structure which will be briefly explained in this section. Usually, an essay starts with an introduction which includes a short description of the topic. The introduction describes the controversial topic of the essay and rarely includes arguments. However, the introduction frequently include a *thesis statement* which expresses the stance of the author about the topic. We refer to this statement as *major claim* (cp. section 2.2).

The actual arguments which either support or attack the major claim are given in the paragraphs following the introduction. Usually there are about two or four paragraphs before the essay concludes with a concluding paragraph. The last paragraph frequently includes a re-statement of the major claim which we will also annotate. It might also include a summary of the reasons supporting the major claim which we will annotate as claims either for or against the major claim. In rare cases the last paragraph also includes complete arguments which should also be annotated. Frequently, the very last sentences include some *recommendations* for future actions which can be considered as a result of the authors' discussion. These recommendations are not argumentative and should be neither annotated as major claim nor as arguments.

Figure 1.3 illustrates the common structure of persuasive essays including the introduction, the body paragraphs and the concluding paragraph. The actually argumentation structure which we aim to annotate in this study will be a tree structure. The root node of this tree is the major claim (if the major claim is restated several times in the introduction or in the conclusion, this node will include

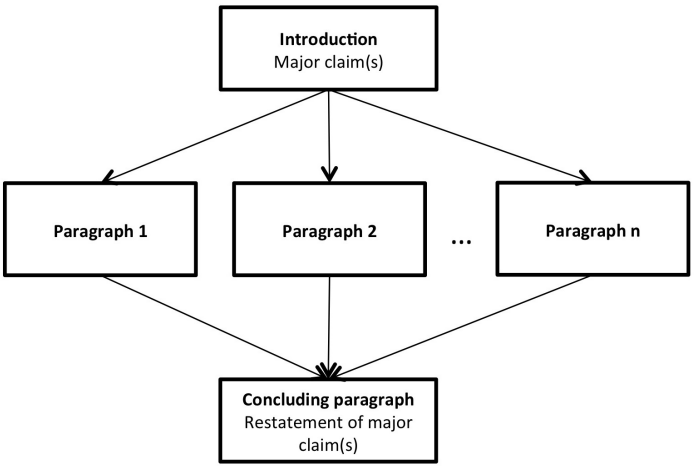

Figure 1.3: Common structure of persuasive essays

several but semantically very similar statements/annotations). The following nodes in the tree structures are the claims of the arguments and the premises are the reasons given for underpinning the claims. To illustrate this in more detail figure 1.4 illustrates an example argumentation structure. The relations from the claims of

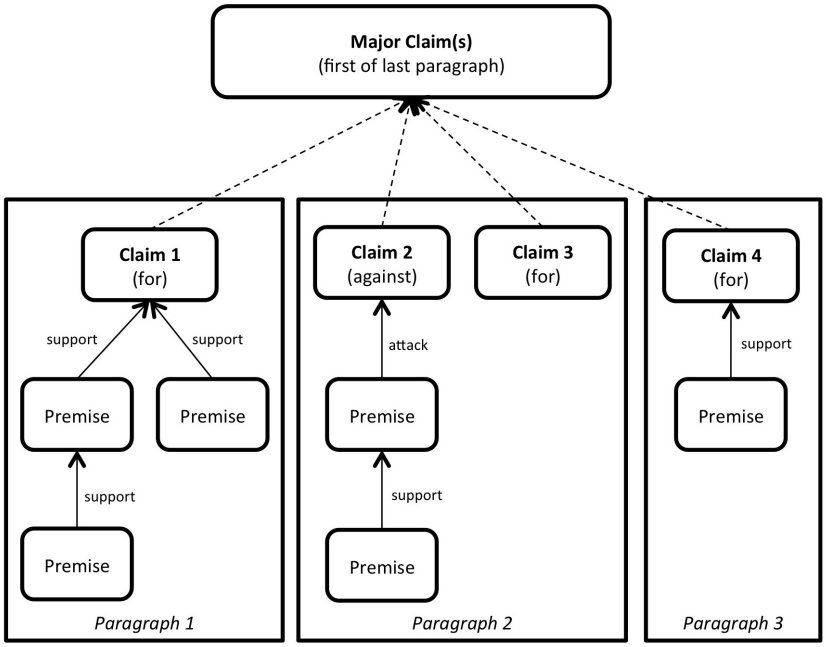

Figure 1.4: An example of the argumentation structure of a persuasive essay

the arguments to the major claim are dotted since we will not explicitly annotated them. The relation of each argument to the major claim is indicated by a *stance attribute* of each claim. This attribute can either be *for* or *against* as illustrated in figure 1.4. Having described the argumentation structure that we aim to annotate in this study, the next section provides a brief overview of the annotation process

before a more detailed description of each individual annotation step is given in the following chapters.

1.3 Overview of the Annotation Process

Previous sections briefly described the argumentation structures that we aim to annotate in this study. For annotating these structures, we split the annotation process into two steps: (1) Annotation of argument components and (2) annotation of argumentative relations. Each of these steps is further divided into different smaller sub steps

Annotation of argument components
The first step of the annotation process focuses on the annotation of argument components. It is further divided into 4 steps which should be followed in the described order:

1. Annotation of the Major Claims: In this step we annotate the major claim(s) which are either located in the introduction or conclusion of the essay. In seldom cases it might also be possible that several reformulations are present. In these cases each should be annotated. The details for annotating major claims are given in section2.2.

2. Annotation of Claims: Each claim is the central component of an argument which either supports or attacks the major claim. So, it can be considered as a reason for the major claim. Claims are usually included in the paragraphs between the introduction and conclusion. In some cases there are also claims present in the introduction or conclusion, e.g. as a summarization as the key reasons for the major claim. These should also be annotated. Each claim has a stance attribute which incites if the argument is either for or against the major claim of the author. A detailed description how to annotate claims is provided in section 2.3.

3. Annotation of Premises: In the final step of the argument component annotation, the premises for the claims are annotated. These are the reasons given by the for supporting or attacking the claims. Usually, the premises are included in the body paragraphs located closely to the claims. The detailed description of the annotation of premises is given in section 2.4.

Annotation of argumentative relations
The second step focuses on the annotation of argumentative relations. In particular, we will annotate support and attack relations holding between the argument components to identify the structure of the arguments. For instance, it might be possible that some premises are sequentially connected with support relations or conversantly support a claim. The details of this step are described in chapter 3.

Chapter 2

Annotation of Argument Components

This chapter describes the annotation of argument components in detail. Section 2.1 describes some general rules for annotating argument components. These rules focus on the aspect of the boundaries of argument components which should be followed for each of the three argument components (major claims, claims and premises). Section 2.2 focuses on the annotation of major claims whereas the following two section focus on the the annotation of claims and premises. In particular, the annotation of claims and premises is closely related. So, it is important to read all sections and to gain a good understanding before starting with the actual annotation process. Each section includes several examples explaining the details of the argument component annotations in detail[1].

2.1 Argument Component Boundaries

The argument examples in section 1.1 already showed that argument components do not necessarily cover a whole sentence. Frequently, a sentence might include several argument components which should be annotated separately (e.g. a claim and a premise in one sentence). In addition, so called "*shell language*", e.g. phrases like "*I am strongly convinced*", "*Another reason is that*" or "*From all these reasons follows that*" are not relevant for the content of the argument and should not be annotated. In this section, we will first provide a list of general rules and some examples which should help to identify the boundaries of each argument component.

Completeness Rule: An argument component should always cover a statement, which can stand in isolation as a complete sentence. A simple test to verify if an annotated component is a complete statement is to prepend the clause "*It is true that, <claim>*". If the resulting sentence is grammatically correct, the annotation is valid according to the completeness rule.

[1]Some of the examples are taken from real essays either from `http://www.buowl.boun.edu.tr` or `http://www.essayforum.com`

Relevance Rule: Include all words which are relevant for the argument component. This means that all content relevant subordinate clauses should be included in the component annotation. This also includes temporal information like "*In ancient times*", "*Recently*" or "*These days*" at the beginning of a sentence since it might be not possible to understand the whole argument without these times specifications.

Shell Language Rule: As mentioned before, shell language has no bearing on the context and thus, it is not relevant for the arguments. Indeed such phrases might include indicators which facilitate the identification of argument components however, for the content of the argument they are not important. Besides the above mentioned expressions, shell language might include terms and phases like "*For example*", "*According to the previous fact*", "*As can be seen*", "*Another important point which contributes to my argument is that*", "*I agree to this view that*", "*In this context*", etc. There is only one *exception* in which shell language should be included in the annotation, namely if it contains a negation that is important for the content of the argument. For example, the phrases "I do not agree that", "I disagree with the view that", etc. In these cases the shell expression should be included since it changes the meaning of the argument component.

Splitting Rule: A sentence should only be annotated completely if and only if it does not include an inference step between several statements and if it does not include shell language. That means, even if a sentence includes several complete statements, the sentence should only be splitted into two (or several) argument components if one statement is a reason for the other. In particular, it is important that sentence which include several complete statements connected with conjunctions like "*and*" or "*or*" usually do not include an inference step. This might for example happen if a sentence contains several reasons for a claim expressed in another sentence. In this case, the two premises should be annotated as one argument component. This also holds for conditional sentence (*if...then-statements*) since, those do not include an inference step between a claim and a premise.

Punctuation Rule: Punctuations at the end of an argument component should not be included in the annotation.

The following examples illustrate correct argument component annotations according to these rules (the text in squared brackets illustrates the correct boundary for the argument component annotation):

> **Example 2.1.1**: "[Because of convenience many people drive with their own car]."

In this example the whole sentence should be annotated because otherwise the relevance rule is violated. The term "*convenience*" is important for the statement. its also important to note that the shell expression "*Because of*" has also to be included because otherwise the component would not be a complete statement.

> **Example 2.1.2**: "This is due to the fact that [school uniforms are quite expensive and not every student can afford them]."

In this example the shell language at the beginning of the sentence is not included in the annotation due to the Shell Language Rule. The following two statements could be indeed annotated as independent statements, however, this would violate the splitting rule, since both are reasons given against wearing school uniforms.

> **Example 2.1.3**: "[More advance sport lessons should be provided during primary education], since [the health of students will benefit from regular exercises]."

This sentence includes a claim and a premise. Therefore, it should be splitted into two argument components.

> **Example 2.1.4**: "[I do not agree with the opponents of nuclear power]."

In this example, the shell language should be included included because otherwise the major claim would not be a complete sentence and thus the annotation would violate the Completeness Rule.

> **Example 2.1.5**: "[I disagree with the viewpoint that school uniforms have positive effects]."

In this example the shell expression at the beginning of the sentence includes a negation which is important for the content of the argument. Therefore, it should be included in the annotation.

2.2 Annotation of Major Claims

In persuasive essays the *major claim* represents the stance of the author about the essay topic. It is also called *thesis statement* and frequently indicated by opinion expressions like "*From my point of view...*", "*In my opinion...*", "*I strongly believe that...*", etc. Usually, the major claim is present in the introduction or conclusion of an essay or in both. In the introduction it has the characteristics of a general assertion or an opinion with respect to the topic, whereas in the conclusion the major claim summarizes the argumentation according to the author's stance.
For annotating major claims you should follow the rules for argument component boundaries described in the previous section. Please also note that all occurrences of the major claim should be annotated and that in some cases even the introduction or conclusion might include several reformulations which should be individually annotated.
For getting familiar with this particular type of argument component, we will investigate some examples of introductions and conclusions from real essays (the major claim is in squared brackets and bold-face):

> **Example 2.2.1** (Introduction): "Cloning is creating a genetic copy or replica of cells, tissues, embryos, and genes of an already existing organism. Thanks to advances in new technology, cloning of animals has

> succeeded, but a human has not been cloned so far because of the lack of technology and the prohibition by governments. I think that [**a human cannot be cloned**] because human cloning involves many risks."

In this introduction (2.2.1) the major claim is clearly indicated by the opinion expression "*I think that...*". Following the rules of argument component boundaries not the whole sentence should be annotated since, the opinion expression can be considered as shell language and the statement following the discourse connective "*because*" is a reason given in support for the major claim. Therefore, the sentence includes an inference step between several statements. Since, the second statement in the sentence is directly supporting the major claim it should be annotated as a claim (cp. section 2.3).

> **Example 2.2.2** (Conclusion): "As a result, [**human cloning may have advantages but no disadvantages**]. Yet there is no experiment of cloning a human so we should try it first by cloning every human for possible organ transplantation. Thus, we can make sure that cloning is used for good intentions."

In this conclusion (2.2.2) the major claim is not indicated by an opinion expression. However, it should be obvious that the first sentence includes the summarizing statement which represents the stance of the author. The shell expression "*As a result,*" is not included in the annotation.

> **Example 2.2.3** (Conclusion): "To sum up, for most people it might be the biggest happiness to have children. However, I firmly believe that [**having children is not everything in life**]. People can also live full and accomplished lives without children."

In example 2.2.3 the major claim is again signaled by an opinion expression which is not included since it is not important for the content of the major claim.

> **Example 2.2.4** (Conclusion): "In sum, I think human cloning may cause a lot of big and important problems if humans are cloned. However, I think cloning of animals and organs is beneficial. So [**'yes' to cloning of animals and organs, but 'no' to human cloning**]."

Example 2.2.4 includes several opinion expressions which represent conditional statements. However, the major claim which represents the stance of the author is included in the last sentence.

> **Example 2.2.5** (Introduction): "There are many ways for people to be happy. Some people are happy maintaining successful business affairs, some of them are happy having big amounts of money and many of them are happy bringing up their children. In my opinion, [**children are the ultimate bliss in our lives and if I reach the suitable age for marriage; I really want to have at least two children**]. Not only me but also many people plan to have children of their own as they add beauty to our lives."

In example 2.2.5 the major claim covers several statements. Since there is no inference included in this sentence, all the statements except the shell phrase at the beginning are annotated as one single major claim.

> **Example 2.2.6** (Introduction): The idea of school uniforms seems like an antiquated concept for many people. Unless a child attends private school, it is not normally practiced by children and families. Students studying in schools requiring school uniforms generally perform very well academically and seem happy wearing the same outfit every day. [**There are many benefits to wearing school uniforms that schools all around the world should incorporate into their public schools**].

Example 2.2.6 does not include any indicator like opinion expressions. In this case it is more complicated to identify the major claim. However, from the context ist should be clear that the previous sentence are some reasons supporting the stance of the author, namely the last sentence which is the major claim.

> **Example 2.2.7** (Introduction): "As the way to cloning has been found, there has been a debate about if it is right or wrong to clone a human. Some people think that cloning is beneficial for humankind while other people argue that [**cloning has too many disadvantages**]. I agree with the latter view and will give several reasons in the following paragraphs."

In example 2.2.7 the last sentence includes a stance expression which might signal the presence of a major claim. However, since this sentence does not include any content relevant information, the referenced statement should be annotated.

> **Example 2.2.8** (Introduction): "And finally medicine with the help of technology has developed its most extreme product, the human being! Some say that it is a big step towards immortality. However, I strongly disagree with this view because of the following reasons."

The introduction in 2.2.8 illustrates a complicated example. The author disagrees with a given statement. Just annotating the statement would be wrong because the author's stance is actually the opposite. The last sentence includes nothing content relevant. Therefore, we are not able to annotate a major claim. If the last to sentence would be incorporated in one single sentence we could annotate the whole sentence. However, these annotation guidelines does not allow to annotate several sentence as an argument component. So, we should search the major claim in the conclusion and if it is also not present, we will not annotate a major claim.

Note that in some essays the introduction or conclusion also includes reasons which support the major claim. In these cases it might be complicated to distinguish between major claims and the supporting statements which should be annotated as claims (cp. section 2.3). Example 2.2.9 illustrates such a case:

> **Example 2.2.9** (Introduction): "... . I believe that [**we should invest in cloning technology**]. Cloning helps to develop new cures for lethal diseases. Cloning of animals and organs can be beneficial for humankind. It could also foster research in biomedicine which would be an important step towards new technologies."

In this example only the first sentence contains the major claim, since it represents the stance of the author towards the topic (in this case "*Cloning*"). The second, third and fourth sentences provide several reasons to support the author's stance. Therefore these sentences (2-4) do not include major claims but reasons which should be annotated as claims.

2.3 Annotation of Claims

A claim is a direct support (or refutation) of the author's stance. So, it is a direct reason given in support (or attack in the case of a contra argument) of the major claim. In body paragraphs[2], a claim is usually supported with one or several reasons/premises whereas in the introduction or conclusion, a claim appears as a direct reason of the major claim. Commonly, the claim is an assumption that should not be accepted without additional support. Since the characteristic of claims in body paragraphs might differ from claims in introductions or conclusions, we distinguish these two cases in our guideline. However, in both cases, each claim has a *stance attribute* which denotes if the claim is "*for*" or "*against*" the major claim. We will illustrate this attribute in the examples given in the following sections.

2.3.1 Claims in Body Paragraphs

A claim in an body paragraph is the central component of an argument. It appears frequently as an *initial assumption* located at the beginning of a paragraph or as a *conclusion* near the end. In few cases, the claim might also be located somewhere between the statements of a paragraph. Most frequently, one paragraph includes a single unique claim and there are only few cases where several claims/arguments are included in a single body paragraph. In this case, a paragraph includes several arguments covering different topics or aspects related to the topic.
Commonly, a claim does not appear without reasons (premises) in an body paragraph. So, for annotating claims in body paragraphs, it might help to identify which statements are reasons for others. If there is one statement which is not a reason for another statement in the paragraph, it is likely the claim of an argument. During the annotation process it might also help to be aware of the major claim, since the claims are direct reasons for the stance of the author. So, it is likely that claims in body paragraphs share some entities with the major claim (e.g. locations, persons or general noun phrases).
For getting familiar with the annotation of claims in body paragraphs, we will investigate some examples. In each example, the major claim of the essay is included

[2] paragraphs between introductions and conclusion

since it might help to identify the claims in body paragraphs. In each example the claim/s of the paragraph is/are in square brackets and underlined.

> **Example 2.3.1.1**:
> Major Claim: "Cloning is an important technology for humankind"
> Paragraph: "[The technology of cloning can be helpful for developing new cures]. For example, by reproducing organs like kidneys or levers, many people with serious diseases can be healed. In addition, it might be helpful for understanding other important processes in the area of gene technology which can help to invent new kinds of treatments."

In this example, the claim is given as the first statement in the paragraph. It is an initial assumption and a direct support of the major claim. The two following sentences are reasons given for the claim. These reasons are the premises which support the claim. Since the first statement does not serve as a reason for another statement (there are no outgoing support relations, cp. section 3), it is likely to be the claim. Also note that the rules for argument component boundaries hold for claims. The stance attribute of the claim in this example should be set to "*for*" since it supports the major claim.

> **Example 2.3.1.2**:
> Major Claim: "It is dangerous to clone humans"
> Paragraph: "The consequences of cloning humans are incalculable since no scientist has ever cloned a human being. It might cause terrible consequences and uncontrollable changes in the human gene pool. Therefore, [human cloning should be prohibited]."

In this example, there are two candidates which could serve as a premise. The first part of the first sentence seems to be a direct support for the major claim. However, this statement is a support for the last sentence, so it cannot be the claim of this paragraph. The last sentence however starts with the indicator "*therefore*" which is a strong signal for the presence of a claim. In addition, the first statement is a reason given why human cloning should be prohibited. The second sentence also seems to be a good reason for this statement. So the last sentence does not serve as a reason for another statement in this paragraph; it seems to be the conclusion of all the reasons given and therefore it should be annotated as the claim. The stance of this claim is again "*for*" because it supports the major claim.

> **Example 2.3.1.3**:
> Major Claim: "Some museums will not disappear"
> Paragraph: "Admittedly, [it is more convenient to learn about historical or art items online]. With Internet, people do not need to travel long distances to have a real look at a painting or a sculpture, which probably takes a lot of time and travel fees."

The third example includes a contra argument against the major claim. So the stance of the claim in the first sentence is set to "*against*". The second statement supports this contra claim; therefore, it is a premise given and the first statement is the claim of this argument.

Example 2.3.1.4:
Major Claim: "Technology negatively influences the way how people communicate"
Paragraph: "[Some people use their cellphone everywhere and do not even notice their environment]. Furthermore, [the language changed due to this new technology]."

Example 2.3.1.4 is a very infrequent case. There are two reasons given for the major claim. The first is about the overuse of cell phones and the second about the influence of technology on language. These two topics are not related and neither does the first statement support the second nor the second the first. However, since both are reasons for the major claim, the two statements are annotated as claims. The stance attribute of both claims should be set to "*for*" since both support the major claim. Note again that this case, occurs relatively infrequently in persuasive essays.

Example 2.3.1.5:
Major Claim: "School uniforms should be mandatory in each school."
Paragraph: "First, by wearing the same clothes, students learn to judge people without looking at their appearance and expensive brands of clothing. [School uniforms will decrease bullying which is very common in todays' schools]. If someone looks richer, most people feel like they have a higher social status or more power."

In this example the claim is located in the middle of the paragraph. It is a direct reason given for the major claim and the first and last sentence include some reasons which support the claim. Also note, that in this example no indicators are present. For identifying the claim in this case, it is necessary to recognize how the statements support each other. The stance is again "*for*".

Example 2.3.1.6:
Major Claim: "Cloning is a new technology that is necessary for our world"
Paragraph: "First, [cloning organs is useful for the treatment of lethal diseases]. Thanks to organ transplantation by cloning, people may be healthier and happier. Furthermore, [cloning animals enables developments in science] because more animals can be used for experiments. This is the useful side of cloning."

This example includes two different arguments. This is denoted by two different aspects of the topic. The first argument is about healing diseases by using cloning and the second aspect is about novel (more general) developments in science. So this example includes two different arguments which both support the major claim. Therefore, the stance attribute of each claim is "*for*".

Example 2.3.1.7:
Major Claim: "Cloning is a threat for our society."

> Paragraph: "By cloning humans there would be a split in the society. Clones would have extraordinary abilities which are a result of improving their human genes. For instance, genes could be manipulated in such a way that this novel human generation is immune against common diseases. Consequently, [it would be difficult for clones to integrate well in today's society]."

In this example, the first and the last sentence seem to be good candidates for the claim. Both are similar, the given reasons support both statements and both seem to be good reasons for the major claim. However, on closer inspection the statement in the first sentence emerges as a reason for the last statement. In addition, the last statement includes the indicator "*consequently*" which is a strong indicator for a conclusion. Therefore, the last statement should be annotated as a claim. It's stance is "*for*" since it supports the major claim.

Example 2.3.1.2 and example 2.3.1.7 showed that indicators like discourse connectives can facilitate the identification of claims. Appendix A includes a list of claim indicators which frequently signal the presence of claims. However, these indicators are not a warrantor for the presence of a claim. There are some cases in which such an indicator is used for a preliminary 'result' e.g. in a reasoning chain. So, even if one or several of these indicators is/are present in a paragraph, it is necessary to understand the content and to recognize which statements support (or attack) each other. As mentioned above, the major claim should be also considered when searching for a claim in paragraph.

2.3.2 Claims in Introductions and Conclusions

In the introduction or conclusion, a claim appears as a direct reason (or refutation) of the major claim. In many cases, the claims are located adjacent to the major claim. In contrast to body paragraphs, the introduction or conclusion infrequently includes complete arguments including premises and often the claims are reformulations of the arguments included in body paragraphs (for instance it is likely that a conclusion contains a condensed version of the key claims given previously). The following examples illustrate claims in introductions and conclusions. In these examples the major claim is in bold face (if present) and the claims are underlined. Both are also put in square brackets to illustrate the boundaries more precisely.

> **Example 2.3.2.1** (Introduction): "Do you want a twin that is cloned from you? Do you think it is necessary? Or do you think it is unethical and should be banned? I strongly believe that [**cloning is a new technology that is necessary for our world**]. I have various reasons for this: [it is necessary for the treatment of some illnesses such as leukemia and it provides our children better lives]."

This example starts with three rhetorical questions which introduce and clarify the debatable character of the topic. The following sentence includes a very precise standpoint of the author followed by reasons given as support for it. These reasons

directly support the major claim. Therefore, they are annotated as claims with the stance attribute set to "*for*". Since the last sentence includes an enumeration of several direct reasons for the major claim it is not split into several argument components.

> **Example 2.3.2.2** (Conclusion): "To sum up, although [cloning humans might bear some risks], I strongly believe that [**this technology is beneficial for humanity**]. [It is likely that this technology bears some important cures which will significantly improve the life conditions]."

This example of a conclusion includes two claims surrounding the major claim. The first claim is a rebuttal which illustrates some risks of cloning. It is a statement which directly attacks the major claim or the author's stance respectively, so the stance attribute for this claim should be set to "*against*". The last sentence includes a reason which supports the major claim. Therefore, it is annotated as a claim and the stance attribute of it is set to "*for*".

> **Example 2.3.2.3** (Introduction): "As the way to cloning has been found, there has been an argument about if it is right or wrong to clone a human. Some people argue for and others against cloning, but we cannot reach an agreement because there is no evidence supporting either side. However, as far as I am concerned, [**the disadvantages of cloning outweigh the advantages**] because [the consequence of these experiments are not foreseeable]."

The third example illustrates an introduction where the last sentence includes the major claim. In the same sentence there is a reason given which supports the major claim. Since the reason can be separated as a statement (cp. rules for argumentation boundaries in section 2.1), it is annotate as a claim. Its stance attribute is set to "*for*".

> **Example 2.3.2.4** (Introduction): "And finally medicine with the help of technology has developed its most extreme product, the human being! Some say that it is a big step towards immortality, while some claim that it is something unnatural. Since [it us unnatural and unethical], [**I definitely disagree with the idea of human cloning**]."

In this example, there a reason precedes the major claim. Since, it can be separated using the guideline for argument component boundaries, it is annotated as a supporting claim for the major claim. It's stance attribute is set to "*for*".

> **Example 2.3.2.5** (Conclusion): "To conclude, although [school uniforms are expensive], I think [**they should me mandatory in each school**]. [They will prevent bullying] because by wearing the same clothes pupils do not judge their classmates by their appearance and brand-name cloths."

The last example illustrates an argument following the major claim in a conclusion. The last sentences includes two statements. The first statement is a direct reason given for the major claim. Therefore, it is annotated as a claim (stance attribute is "*for*") followed by a premise in the same sentence. In addition, a contra claim precedes the major claim. It is a contra reason and therefore its stance attribute is set to "*against*".

2.4 Annotation of Premises

In this step, we focus on the annotation of the second argument component: the premise. A premise is a reason given for supporting or attacking an argument component. So it can be considered as a justification or refutation for convincing the reader of the truth or falsity of a claim. This also means that a premise is always connected to another argument component which could either be a claim or another premise if there are reasoning chains included in a paragraph. Note that in contrast to the claim annotation, the stance of each premise is not encoded in an attribute. It will be annotated by support or attack relations which link the premises to claims or other premises. For example the statement "*Children bring happiness and meaning to your life*" is a supporting premise whereas "*It is a heavy psychological burden to have children*" is an attacking premise for the claim "*Having children is the ultimate bliss in our lives*". In this case both should be annotated as premises, the distinction between supporting and attacking premises by means of argumentative relations is described in chapter 3.

Since the context and the identified claims from the previous step are important for annotating premises, the annotator should search for each claim in a paragraph and find the reasons given for it. It is possible that a claim and a premise are included in a single sentence or that a premise is only a part of a sentence. So the annotation of premises is also conducted at the clause level and the rules for argument components should be followed.

Usually, there are several premises given for a single claim in a paragraph. Sometimes there are also reasoning chains where several premises are linked together for supporting a claim. The following examples illustrate the annotation of premises in detail (the premises are wavy underlined).

> **Example 2.4.1** (Body paragraph): "First, [cloning can help human families to gain children]. For instance, [parents with no eggs or sperms can create children that are genetically related]. [Even same sex couples can have children without the use of donor sperm or donor eggs]."

In this example, the first statement is a claim which is supported by two premises. Both of the premises are indented to convince the reader that cloning is a positive development and that there are particular families which will benefit from this technology. Note that this example includes two argumentative relations connecting the two premises to the claim. Both of these relations indicate that the source argument components (the two premises) are reasons given for the target component (the claim). Having this structure in mind also helps to identify the claim. As mentioned in the introduction of this guideline, the argumentation structure is always

a tree and the root node of a tree is always a claim. So, for annotating argument components, it is helpful to imaging the argumentation structure and to recognize which statement supports or attacks another one and vice versa. In this example, it is only possible to connect the argument component two and three to the first one. So the first component, which is the root of the argumentation structure is likely to be the claim of the paragraph.

> **Example 2.4.2** (Body paragraph): "[There would be a lack of uniqueness and violate convictions regarding human individuality and freedom]. So, [clones could be seen as less than human compared with non-clones]. Therefore, [human cloning would divide our society into two different groups]."

This example illustrates a reasoning chain including one claim at the end of the paragraph and two preceding premises. The first premise is a reason given for the second premises which is a reason given for the claim. The chain of reasoning is indicated by indicators. Both indicators, "*so*" and "*therefore*", signal a conclusion based on preceding statement(s). However, since all statements form a chain, only the last statement (or the root of the tree) is annotated as a claim.

> **Example 2.4.3** (Body paragraph): "[Having children is the ultimate bliss in our lives]. There are many reasons which support my viewpoint. For example, [a close friend of mine is so happy since her child was born]. [She says that raising a child is like having an important goal in life] since [bringing up a human being is more satisfactory than anything else]."

This example illustrates a paragraph where a sentence does not include argumentative content. The second sentence only states that there are reasons without stating them. So, it is not annotated as a premise. The third sentence includes an example referring to a close friend. It is some kind of evidence indented to support the claim in the first sentence. Therefore, it is annotated as premise. The fourth sentence includes two premises which are connected by means of the discourse connective "*since*". Both are annotated as premises because the sentence includes an inference (cp. section 2.1).

> **Example 2.4.4** (Conclusion): "To conclude, although [having children could be considered as a financial burden], I think that [**it is the ultimate goal**]. [Nobody will regret to have children] since [having children brings happiness and meaning to our lives]."

Example 2.4.4 illustrates a conclusion including a premise. Generally, this case is quite infrequent. The conclusion begins with a claim against the major claim followed by another claim supporting the major claim in the last sentence. Since the last sentence includes two statements and the second statement supports the preceding one, the second statement is annotated as a premise for the preceding claim. Note again, that direct reasons given for the major claim should be annotated as claims and not as premises. Only if there are additional reasons given for a claim in a conclusion, like in example 2.4.4, they should be annotated as a premise.

Example 2.4.5
MajorClaim: "Cloning is an important technology for humankind"
Body paragraph: "Some people believe that [human cloning would divide our society into two groups]. They argue that [clones will be seen as less compared to naturally born humans] because [clones will lack uniqueness and individuality]. However, [a human clone will not differ that much from other humans]. [They could develop exactly the same abilities as naturally born humans]."

This example shows a more complicated argumentation structure. The paragraph starts with a contra-claim stating that human cloning has some negative effects on our society. This claim is followed by a reasoning chain of supporting premises in the second sentence. The third sentence is an attacking premise of the claim which is supported by the last statement in the last sentence. Note again that the determination between supporting and attacking premises is conducted by means of argumentative relations. In this example the premise "*a human clone will not differ that much from other humans*" will be linked with an attack relation to the claim. The author also adds an additional support to this premise in the last sentence which will be linked using a support relation.

Example 2.4.6 (Body paragraph): "If human cloning became possible what would be the outcome of it? Basically I think [human cloning is against the laws of nature]. [There is a well organized balance of nature and human cloning may damage this wonderful balance] since [the number of people would increase due to cloning]."

Example 2.4.6 begins with a rhetorical question. Generally, questions are non-argumentative since the answer to the question is not known. So it is not annotated as an argument component. The second sentence includes the claim of this paragraph. It is supported by the first statement in sentence 3. Since the last sentence includes an inference and two statements, both are annotated as a premise.

As for claims, premises are sometimes signaled by indicators which facilitate the recognition. For instance, indicators like "*because*", "*the reason for this is*", "*in addition*" may signal the presence of a premise. A list of indicators for premises is provided in appendix B. Note again, that these indicators are not a guarantor for the presence of premises. It is necessary to understand the complete argumentation structure and to recognize which component supports or attack another one. For instance there might be cases where in a reasoning chain a claim indicator is used by the author for emphasizing an "intermediate result". Such an example is shown in 2.4.2. The indicator "*So*" at the beginning of the second sentence might indicate a claim. However, since the argument component following this indicator is only an intermediate result in a reasoning chain this statement is not annotated as the claim of this paragraph.

A particular characteristics of indicators is that they signal the direction of reasoning. For instance, indicators like "*because*", "*since*", "*in addition*", "*for example*"

or "*first*"[3] signal that the following statement refers to a preceding statement whereas indicators like "*therefore*", "*as a result*", "*hence*" or "*thus*" signal that the following statement is a "result" of preceding statements. So, if an indicator is present in a paragraph which signals that the following argument component i is a "result" of preceding statements, the statement i is a good claim candidate. The same holds for indicators which signal that a statement refers to a preceding component. In this case the signaled component is likely to be a premise and not a claim. Recognizing the direction of the reasoning in a paragraph by means of indicators will strongly facilitate the annotation of argument components. So, before annotating any argument component in a paragraph, it might help recognize the direction of reasoning by means of the indicators (e.g. discourse connectives or shell language).

[3]Indicators which signal enumerations like "*first*", "*second*", "*in addition*", "*furthermore*", etc. indeed indicate that the following statement refers to a preceding one. However, frequently these indicators are also used to enumerate the arguments in an essay and therefore the claims. In this case, a paragraph might start which an enumeration indicator followed by a direct reason for the major claim which is frequently the claim of a paragraph. If enumeration indicators are present inside of an paragraph it is likely that they enumerate the premises for a claim.

Chapter 3

Argumentative Relations

Having annotated the argument components, we link the argument components with argumentative relations in order to build the tree structure of each argument. An argumentative relation is a directed link between two argument components with a particular source and target component. Such a relation either indicates that the source component is a justification (support relation) or a refutation (attack relation) for the target component. Since, argumentative relations between claims and major claims are implicitly encoded in the stance attribute of the claim and we assume that those relations are the only ones which cross paragraph boundaries, we focus on the annotation of argumentative relations in paragraphs only. Therefore, each source component of an argumentative relation is a premise. However, the target is not restricted to be a claim, since there might be deeper tree structures including serial support. In other words, by annotating argumentative relations we identify for each premise the target it belongs to and recognize if the premise supports or attacks the target. Note that the target can either be a claim or another premise.

The following process illustrates the annotation process. For each paragraph including claims and premises the argumentation structure is build using the following steps:

1. Select a claim c
2. Link each premises in the paragraph if it obviously supports or attacks the claim.
3. For all not connected premises in the paragraph, test if it could be connected to an already connected premise. If that is not possible reformulate the premise and connect it to a matching claim or premise in the same paragraph

Following this process, ensures that the argumentation structure is a tree and that each premise is linked to exactly one argument component (either claim or premise). Also note, that in some cases, the support (or attack) of a single premise might be weak and not obvious. However, when combined with another premise the reason might become stronger. This happens for example if the author uses a particular event or an example in order to justify or refute a standpoint. So sometimes it might be necessary to consider premises in combination in order to identify the correct target of each individual premise. In the next sections, we will illustrate the annotation of support an attack relations in detail by providing several examples.

3.1 Annotating Support Relations

A support relation between two argument components indicates that the source component is a reason or a justification of the target relation. In the following examples, the claim is underlined, the premise is wavy underlined and the argument components are enumerated using superscripts. Let's consider the following simple example:

> **Example 3.1.1** (Simple Example): "[An advanced gun background check should become routine in all gun sales]$_1$ because [it will prevent gun rampages]$_2$."

In this example there are only two argument components of which the first one is the claim and the second one is the premise. The indicator "*because*" signals that the second component is a justification for the first one. Therefore, the second argument component should be linked to the first one using a support relation.

> **Example 3.1.2** (Body paragraph): "First of all, [people cannot predict their own future or know what will happen tomorrow]$_1$. [The world is full of disasters such as wars, pollution, famine, drought, starvation, natural disasters and diseases]$_2$. So [it is just a big mistake to have children]$_3$."

Example 3.1.2 includes three argument components. The claim of this paragraph is present in the last sentence. Following our procedure, we first check if the first premise is a reason for the claim. Since, the statement "no one can predict the own future" might be a reason for not having children, the first premise is linked with a support relation to the claim. The second premise seems also to be reason for the claim. Therefore, it is also linked using a support relation.

> **Example 3.1.3** (Body paragraph): "Furthermore, [it is a very heavy psychological and physical burden to have children]$_1$. [A mother carries her baby in her womb for nine months and 10 days and then the baby torments her during and after the birth]$_2$. [There is no peace, no silence or no sleep at home]$_3$. On the other hand, [the father has to work hard and earn more money]$_4$ because [the baby comes with his expenses]$_5$."

Example 3.1.3 includes five argument components. The first components is the claim. Argument component 2 which is a premise seems to be a reason, thus it is linked to the claim with a support relation. Also the argument component 3 is a reason which supports the claim. It is also linked to the claim using a support relation. Identifying the target for the argument component 4 and 5 is more complex. Argument component 4 can be considered as a reason for the claim. Although, argument component 5 seems to be a reason, it should be linked to argument component 4, since both 4 and 5 are included in the same sentence and the author indicates that argument component 5 is a reason for argument component 4 by using the indicator "because". The complete structure is illustrated in figure 3.1.

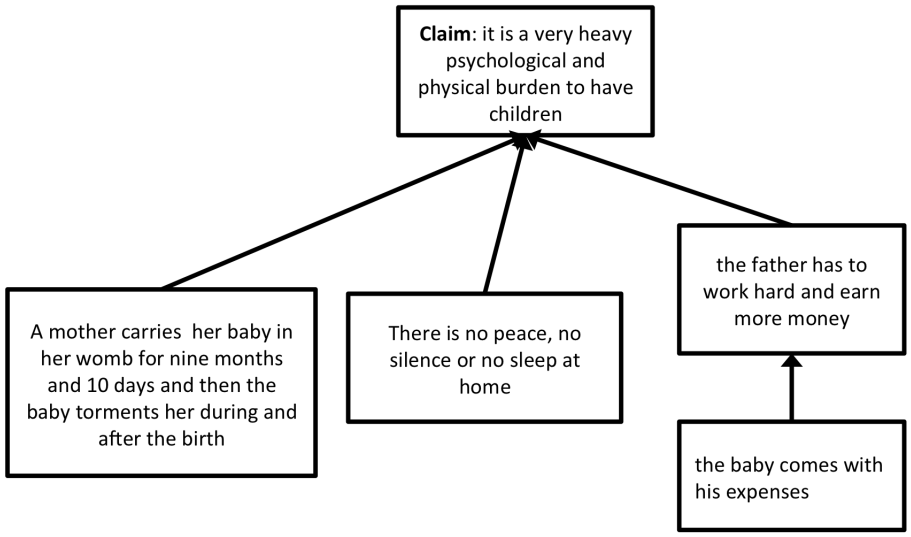

Figure 3.1: Argumentation Structure of Example 3.1.3. All edges are support relations.

> **Example 3.1.4** (Body paragraph): "[Having children is the ultimate bliss in our lives]$_1$. There are many reasons which support my viewpoint. For example, [a close friend of mine is so happy since her child was born]$_2$. [She says that raising a child is like having an important goal in life]$_3$ since [bringing up a human being is more satisfactory than anything else]$_4$."

In example 3.1.4 the first argument component is a claim. Following our procedure, we first check if the first premise is a support for our claim and link it though it is only a weak support. However, the author uses the example of a close friend as evidence it is fine to consider it as a support. Argument component also seems to be a reason for the claim. Therefore it is also linked to the claim. Since argument component 4 is signaled by an indicator to be a support for argument component 3, it is linked to 3 not to the claim though it seems to be a good reason. However, by using the indicator the author explicitly linked it to argument component 3 and not the claim.

> **Example 3.1.5** (Body paragraph): "[Humankind will benefit from cloning]$_1$. For example, [cloning technology can be used to clone organs]$_2$. [It allows to raise new kidneys, levers and other vital organs]$_3$. Thus, [patients in need of organ transplantation will definitely benefit from cloning technology]$_4$. In addition, [cloning can help human families to gain children]$_5$. [It will allow parents with no eggs or sperms to create children that are genetically related]$_6$."

Example 3.1.5 is a more complex examples including six argument components and one claim. The first argument component is the claim in this paragraph. It is a general statement about cloning. The remaining argument components are premises

and there are two different aspects included which are represented in two different branches of the argumentation tree. The first, brach is about cloning for raising organs whereas the second is about cloning for create children. In this example, it is difficult to follow the process described above since, it includes a relatively deep argumentation structure. So we will first separate the two aspects and branches of the tree respectively. The first aspect includes premises 2, 3, and 4. All are about cloning in order to raise organs. And this aspect support the claim of this paragraph in argument component 1. If we consider those three argument components in isolation, it seems that the component 4 is a subclaim since it is signaled with "thus" which indicates a partial result of the reasoning. Therefore, we consider it as the root node of the first branch although 2 and 4 seem to be reasons for the claim in 1. However, since there are two separated aspects which are given by the author for the author, we will model both as different branches in our tree. Argument components 5 and 6 represent the second branch and aspect respectively. Argument component 6 is given as a reason for argument component 5. So we link argument component 5 to the claim and component 6 to argument component 5. The resulting structure is illustrated in figure 3.2

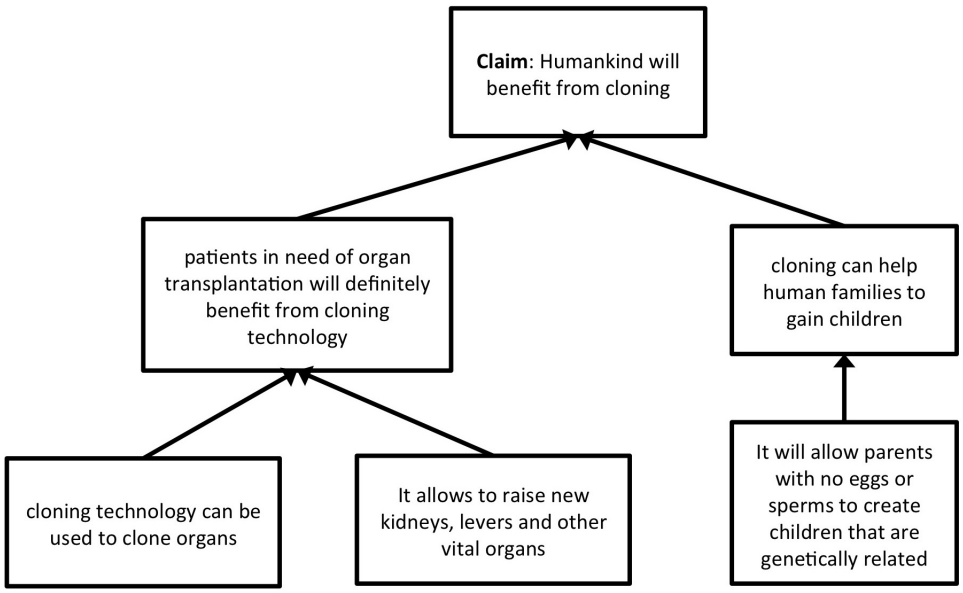

Figure 3.2: Argumentation Structure of Example 3.1.5. All edges are support relations.

> **Example 3.1.6** (Body paragraph): "First, [cloning organs is useful for the treatment of lethal diseases]$_1$. [Thanks to organ transplantation by cloning, people may be healthier and happier]$_2$. Furthermore, [cloning animals enables developments in science]$_3$ because [more animals can be used for experiments]$_4$."

In example 3.1.6 two claims are present. The first one (argument component 1) is about cloning organs for developing novel cures and the second one (argument

component 3) is about cloning animals. Following our procedure for argumentative relation identification, we select one of the two claims. Let's start with argument component 1. Obviously, the premise in argument component 2 is a good support and thus we link it to the first claim. Since the second premise in argument component 4 does not support the first claim, we continue with the second claim. The only premise which is not linked (argument component 4) seems to be a good support. Therefore, we link 4 to the claim in argument component 3.

3.2 Annotating Attack Relations

An attack relation between two argument components indicates that the source component is a refutation or a rebuttal of the target relation. Analog to the previous section, the claim is underlined, the premise is wavy underlined and the argument components are enumerated using superscripts. Let's consider the following simple example:

> **Example 3.2.1** (Simple Example): "[Having children is an incredible experience which everybody should do]$_1$. However, [it also comes along with a lot of responsibilities]$_2$."

This example illustrates a simple case of a attack relations. The claim in argument component is refuted by the premise present in the second component. Therefore, the second component should be linked to the claim using an attack relation. In this example the attack relation is signaled by the indicator "*However*" preceding the premise in the second argument component.

> **Example 3.2.2** (Body paragraph): "[Raising your own child is like having an important goal in your live]$_1$. Admittedly, [you will have great responsibilities and you also will have sleepiness nights]$_2$ but [these drawbacks will turn into a valuable experience when your kids become older]$_3$. Therefore, [Having children is the ultimate bliss in our lives]$_4$."

Example 3.2.2 includes 4 argument components of which the last one is the claim stating that having children is the ultimate bliss in our lives. The first argument component includes a reason for supporting this standpoint followed by a refutation or a doubt in argument component 2 which states that having children is related to having great responsibilities which is a negative point of having children. Therefore, argument component 2 is linked with an attack relation to the claim. In the same sentence, the author refutes this premise again which is signaled by the indicator "*but*". Therefore, argument component 3 is linked with an attack relation to argument component 2. Note that this constellation is a common practice to prevent any potential criticism. By including in your argument a potential rebuttal (argument component 2) and stating why it is not relevant (argument component 3), the overall argument becomes stronger. Also note, that in this example, there are two good candidates for the claim, the first and the last component. However, since the last component is more general, signaled with the indicator "*therefore*", and the first argument component is a reason given for the last component, the last argument

component is annotated as the claim. The structure of this argument is illustrated in figure 3.3.

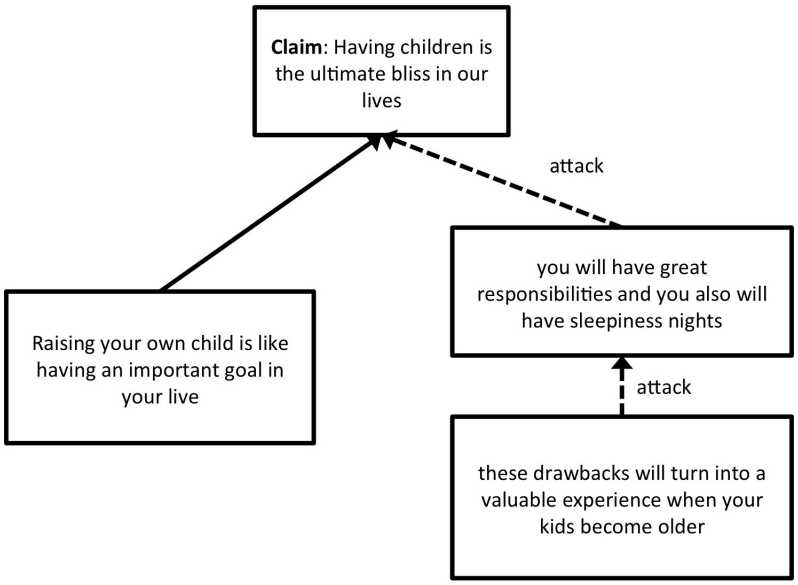

Figure 3.3: Argumentation Structure of Example 3.2.2. Dashed arrows indicate attack relations and solid lines support relations.

Example 3.2.3
MajorClaim: "Cloning is an important technology for humankind"
Body paragraph: "Some people believe that [human cloning would divide our society into two groups]$_1$. They argue that [clones will be seen as less compared to naturally born humans]$_2$ because [clones will lack uniqueness and individuality]$_3$. However, [a human clone will not differ that much from other humans]$_4$. [They could develop exactly the same abilities as naturally born humans]$_5$."

Example 3.2.3 illustrates a common example of an opposing paragraph which might be the last body paragraph of an essay. It starts with a claim against the standpoint of the author followed by a supporting reason (argument component 2) of this contra position. This reason is again supported in the same sentence by the next argument component (argument component 3). In argument component 4 the author starts to attack the contra claim by putting forward a contra reason in argument component 4. Since this contra reason seems to be a refutation of the contra claim in argument component 1, we link it using an attack relation. In argument component 5, the author provides another reason why his reason is true. So argument component 5 is linked to argument component 4 with an support relation. The whole structure of this argument is illustrated in figure 3.4

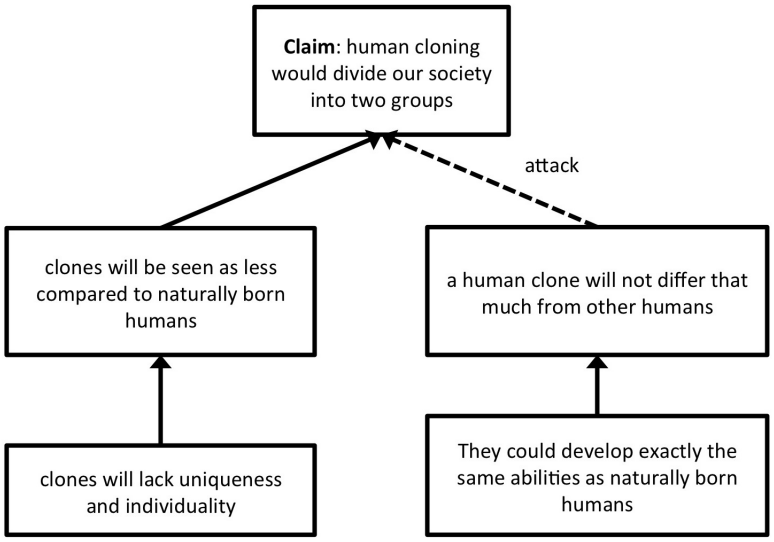

Figure 3.4: Argumentation Structure of Example 3.2.3. Dashed arrows indicate attack relations and solid lines support relations.

> **Example 3.2.4** (Conclusion): "To sum up, although [human cloning bears some risks for our society]$_1$, [the benefits of human cloning still outweigh its drawbacks]$_2$. Therefore, I strongly believe, that [**human cloning should be allowed in order to improve our medical system**]$_3$."

Example 3.2.4 is a conclusion including a major claim in argument component 3, a contra claim in argument component 1 and a contra premise in component 2. Since the first argument component directly attacks the major claim in argument component 3, it is annotated as a claim with its stance attribute set to "*against*". The following argument component is a reason given that the preceding contra claim is not true. Therefore, it is linked to to argument component 1 using an attack relation.

Bibliography

Philippe Besnard and Anthony Hunter: *Elements of Argumentation*, MIT Press, 2008.

M. Anne Britt and Aaron A. Larson: 'Constructing representations of arguments', *Journal of Memory and Language* 48 (4): 794 – 810, 2003.

Raquel Mochales-Palau and Marie-Francine Moens: 'Argumentation Mining: The Detection, Classification and Structure of Arguments in Text', in: *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ICAIL '09, pp. 98–107, ACM, New York, NY, USA, 2009, Online: `http://doi.acm.org/10.1145/1568234.1568246`.

Andreas Peldszus and Manfred Stede: 'From Argument Diagrams to Argumentation Mining in Texts: A Survey', *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)* 7 (1): 1–31, 2013.

Chris Reed, Raquel Mochales-Palau, Glenn Rowe, and Marie-Francine Moens: 'Language Resources for Studying Argument', in: *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, LREC '08, pp. 2613–2618, Marrakech, Morocco, 2008.

Stephen E. Toulmin: *The uses of Argument*, Cambridge University Press, 1958.

Appendix A

Claim Indicators

accordingly
as a result
consequently
conclude that
clearly
demonstrates that
entails
follows that
hence
implies
in short
in conclusion
indicates that
it follows that
it is highly probable that
it should be clear that
it should be clear
points to the conclusions
proves that
shows that
so
suggests that
the point I'm trying to make
therefore
thus
to sum up
we may deduce

Appendix B

Premise Indicators

assuming that
as
as indicated by
as shown
besides
because
deduced
derived from
due to
for
for example
for instance
for the reason that
furthermore
given that
in addition
in light of
in that
in view of
indicated by
is supported by
may be inferred
moreover
researchers found that
this can be seen from
since
since the evidence is
what's more
whereas