BACHELOR THESIS
COMPUTER SCIENCE

RADBOUD UNIVERSITY

Herbrand's theorem

Author:
E.J. Gerritse
s4248120

First supervisor/assessor:
dr. F. Wiedijk
`freek@cs.ru.nl`

Second supervisor:
prof. dr. J.H. Geuvers
`h.geuvers@cs.ru.nl`

April 1, 2016

Contents

1 Introduction

Skolemisation is a way of transforming formulas such that they contain only one quantifier. This makes it easier to prove them using an automated theorem prover (ATP). However, when you prove the Skolemised formula using an ATP, you will not get a proof for the original formula. The rewriting from the proof of the Skolemised formula to a proof of the original formula is called deskolemisation.
Similar to Skolemisation, we have Herbrandisation. These two concepts are dual to each other. In this thesis we are going to focus on Herbrandisation. We are going to look at the lengthy proof of this, together with some examples for clarification.
The proof of Herbrand's theorem is given in Shoenfield [1]. However, Shoenfield is a complex book with a lot of information irrelevant to proving Herbrand's theorem. There are also almost no examples in the book. Thus the focus in this thesis was finding all the relevant theorems and definitions to proving Herbrand's theorem, and giving some examples to make it easier to understand.
About Jacques Herbrand:

Jacques Herbrand (12 February 1908 – 27 July 1931) was a French mathematician who worked in mathematical logic and class field theory. Herbrand's theorem refers to two completely different theorems. One is important in class field theory, the other one will be the Herbrand's theorem which this thesis is about. Herbrand found these results in his doctoral thesis.
Herbrand died when he was 23 when he fell of the Massif des Écrins while hiking with his friends. Even though his death at young age, he was he was already considered one of "the greatest mathematicians of the younger generation" by his professors Helmut Hasse and Richard Courant.

2 Basic definitions and rules

In this chapter we will define the basics for first order languages. How is a first order language constructed, what are the axioms and the rules that we use? It also gives definitions of the basics like theories and theorems.

2.1 Formulas

Definition 2.1. A *first order language* L is a three-tuple of F, R, C:

1. A set of functions $F = \{f_1, f_2, ...\}$, each f_i having an arity $n \in \mathbb{N}$
2. A set of relations $R = \{r_1, r_2, ...\}$, each r_i having an arity $n \in \mathbb{N}$
3. A set of constants $C = \{c_1, c_2, ...\}$

Remark 2.1.1. One can also view constants as 0-ary functions, but in this case we define them as constants. This is to distinguish them from functions more easily.

Definition 2.2 (Terms)**.** We inductively define the *terms* of a language L:

1. A variable $x, y, ...$ is a term.
2. A constant is a term.
3. If $a_1, ..., a_n$ are terms and f is n-ary, then $f(a_1, ..., a_n)$ is a term.

Definition 2.3 (Atomic Formula)**.** We call $r(a_1, ..., a_n)$ an *atomic formula*, in which r is an n-ary relation and $a_1, ..., a_n$ are terms.

Definition 2.4 (Formula)**.** We inductively define the *formulas* of the language L:

1. If a_1 and a_2 are terms, then $a_1 = a_2$ is a formula.
2. An atomic formula is a formula.
3. If A is a formula, then $\neg A$ is a formula.
4. If A and B are formulas, then $A \vee B$ is a formula.
5. If A is a formula, then $\exists x A$ is a formula.

Remark 2.4.1. We can add brackets around A and B when applying these rules to increase readability. We have that $\exists$ and $\neg$ bind stronger than $\vee$, so $\exists x A \vee \neg B$ is to be read as $(\exists x A) \vee (\neg B)$

2.2 Other logical symbols

Shoenfield only defines his rules for $\exists, \vee$ and $\neg$. However, the other logical symbols get used as well. That is why we introduce the following abbreviations, such that one can interpret every formula as one using only $\exists, \vee$ and $\neg$. We introduce the following abbreviations:

- $\forall x A$ is defined as $\neg \exists x \neg A$.

- $A \wedge B$ is defined as $\neg(\neg A \vee \neg B)$.
- $A \to B$ is defined as $\neg A \vee B$.
- $A \leftrightarrow B$ is defined as $(A \to B) \wedge (B \to A)$.

We call $\forall$ and $\exists$ quantifiers.

2.3 More on formulas

Definition 2.5 (Free and bound variables)**.** A variable x in a formula is called *free* if there is no quantifier on that variable. If a variable is not free, it is *bound.*

Definition 2.6 (Closed formula)**.** A formula is *closed* if no variable is free.

Definition 2.7 (Open formula)**.** A formula is *open* if it contains no quantifiers.

Definition 2.8 (Elementary formula)**.** A formula is called *elementary* if it is either an atomic formula or of the form $\exists xA$. For example $\neg A$ is not an elementary formula.

Definition 2.9 (Variant)**.** We get the *variant* of a part $\exists xB$ by replacing $\exists yB[x := y]$, where y is a variable not free in B.

Definition 2.10 (Prenex form)**.** A formula A is in *prenex* form, if it has the form $Q_1x_1, ..., Q_nx_nB$ in which every Q_i is either $\forall$ or $\exists$, and in which B is open. We call $Q_1x_1, ..., Q_1x_n$ the *prefix* and B the *matrix.*

Remark 2.10.1 (Prenex operations)**.** We can use the following prenex operations to get a formula A in prenex form:

1. Replace A by a variant
2. Replace $\neg\forall xB$ in A by $\exists x\neg B$
3. Replace $\neg\exists xB$ in A by $\forall x\neg B$
4. For a quantifier Q, replace $QxB \vee C$ in A by $Qx(B \vee C)$, provided that x is not free in C
5. For a quantifier Q, replace $B \vee QxC$ in A by $Qx(B \vee C)$, provided that x is not free in B

Definition 2.11 (Existential formula)**.** A formula in prenex form is *existential* if all of its quantifiers in its prefix are existential.

2.4 Basic rules

The following section lists the rules we can use when constructing our proofs.

Rule 2.12 ($\exists$-introduction)**.** *If x is not free in B, $A \to B$ implies $\exists xA \to B$.*

Rule 2.13 (Expansion rule)**.** *A implies $B \vee A$.*

Rule 2.14 (Contraction rule)**.** *$A \vee A$ implies A.*

Rule 2.15 (Associative rule)**.** *$A \vee (B \vee C)$ implies $(A \vee B) \vee C$.*

Rule 2.16 (Cut rule)**.** *$A \vee B$ and $\neg A \vee C$ implies $B \vee C$.*

2.5 Axioms, theorems and theories

Axioms are fundamental laws which we accept without any proof. From axioms, we can prove theorems using our rules.

The following four types of axioms are called *logical axioms*

Definition 2.17 (Identity axioms)**.** Axioms of the form $x = x$.

Definition 2.18 (Equality axioms)**.** Axioms of the form $x_1 = y_1 \to ... \to x_n = y_n \to f(x_1, ..., x_n) = f(y_1, ..., y_n)$.

Definition 2.19 (Propositional axiom)**.** $\neg A \vee A$.

Definition 2.20 (Substitution axiom)**.** $A[x := a] \to \exists x A$.

Definition 2.21 (Non-logical axiom)**.** If a theory T (see below) has axioms which are not of one of the four types as defined above, they will be called *non-logical axioms.*

Definition 2.22 (Theory)**.** A *theory* T is a three-tuple of a language, axioms and rules:

1. The language of T, called $L(T)$, is a first order language.
2. The axioms of T are the logical axioms of $L(T)$ and certain further axioms, called the non-logical axioms.
3. The rules of T are the expansion rule, the contraction rule, the associative rule, the cut rule and the $\exists$-introduction rule.

Definition 2.23 (Open Theory)**.** A theory T is *open* if all of its non-logical axioms are open.

Definition 2.24 (Theorem)**.** A theorem A of T is a formula which can be proved by using the axioms and rules of a theory. We then write $\vdash_T A$. If it is clear in which theory we are working, we simply write $\vdash A$.

2.6 Some examples

We will show some examples of proofs using our basic rules and axioms as defined here:

Example 2.25. First we will prove that $B \vee A$ implies $A \vee B$

$$\frac{B \vee A \qquad \neg B \vee B}{A \vee B}\ \text{Cut rule}$$

Since we assume $B \vee A$ and $\neg B \vee B$ is an axiom, the proof is done.

Example 2.26. Next we will show that A implies $\neg\neg A$. If we use something we have proven before, we use a double line as abbreviation.

$$\dfrac{\dfrac{\dfrac{\dfrac{A}{\neg\neg A \vee A}}{A \vee \neg\neg A} \qquad \dfrac{\neg\neg A \vee \neg A}{\neg A \vee \neg\neg A}}{\neg\neg A \vee \neg\neg A}\ \text{Cut}}{\neg\neg A}\ \text{Contraction}$$

Example 2.27. Now let us prove $\neg(A \to B) \to A$. First we have to rewrite this formula such that it only uses $\exists, \vee, \neg$, as seen in section 2.2. This will give us the formula $\neg\neg(\neg A \vee B) \vee A$.

$$
\begin{prooftree}
\AxiomC{$\neg A \vee A$}
\doubleLine
\UnaryInfC{$A \vee \neg A$}
\RightLabel{Contraction}
\UnaryInfC{$B \vee (A \vee \neg A)$}
\RightLabel{Associative}
\UnaryInfC{$(B \vee A) \vee \neg A$}
\doubleLine
\UnaryInfC{$\neg A \vee (B \vee A)$}
\RightLabel{Associative}
\UnaryInfC{$(\neg A \vee B) \vee A$}
\AxiomC{$\neg\neg(\neg A \vee B) \vee \neg(\neg A \vee B)$}
\doubleLine
\UnaryInfC{$\neg(\neg A \vee B) \vee \neg\neg(\neg A \vee B)$}
\RightLabel{Cut}
\BinaryInfC{$A \vee \neg\neg(\neg A \vee B)$}
\doubleLine
\UnaryInfC{$\neg\neg(\neg A \vee B) \vee A$}
\end{prooftree}
$$

2.7 Truth evaluations and tautologies

Definition 2.28 (Truth valuation)**.** A truth valuation v is a mapping from the set of elementary formulas in T to the truth values $\mathbf{T}$ and $\mathbf{F}$ such that for formulas A, B in T.

- $v(A \vee B)$ is true if and only if $v(A)$ is true or $v(B)$ is true.
- $v(\neg A)$ is true if and only if $v(A)$ is false.

Definition 2.29 (Tautological consequence)**.** A formula B is called a *tautological consequence* of $A_1, ..., A_n$ if $v(B) = \mathbf{T}$ for every truth valuation such that $v(A_1) = ... = v(A_n) = \mathbf{T}$.

Definition 2.30 (Instance)**.** A' is an *instance* of A if A' is of the form $A[x_1 := a_1, ..., x_n := a_n]$.

Definition 2.31 (Tautology)**.** A formula A is a *tautology* if $v(A) = \mathbf{T}$ for every truth evaluation v. This means that A is a tautological consequence of the empty set of formulas.

Definition 2.32 (Quasi-tautology)**.** A *quasi-tautology* is a tautological consequence of instances of identity axioms and equality axioms.

Example 2.33. A truth valuation can be seen as a line from a truth table, so for example

A	B	$A \vee B$
0	0	0
0	1	1
1	0	1
1	1	1

Each row from this table represents possible truth evaluation.
$A \vee B$ is a tautological consequence of A, because $v(A \vee B) = \mathbf{T}$ if $v(A) = \mathbf{T}$.

3 Provability

This chapter gives some basic theorems which are used in the proofs in later chapters. The proofs themselves are not given and can be read in chapter 3 and 4.1 of Shoenfield [1]. The reason for this is that most of these proofs are very straightforward and most of the theorems are very standard in mathematical logic.

Assume in this chapter that we are working in a general theory T.

3.1 Definitions and rules

Theorem 3.1 ($\forall$-introduction)**.** *If* $\vdash A \to B$ *and* x *is not free in* A*, then* $\vdash A \to \forall x B$*.*

Definition 3.2 (Closure)**.** Let x_1, ..., x_n be the free variables in a formula A. Then $\forall x_1 ... \forall x_n A$ is called the *closure* of A.

Rule 3.3 (Substitution rule)**.** *If* $\vdash A$ *and* A' *is an instance of* A*, then* $\vdash A'$*.*

Rule 3.4 (Distribution rule)**.** *If* $\vdash A \to B$*, then* $\vdash \exists x A \to \exists x B$ *and* $\vdash \forall x A \to \forall x B$*.*

Rule 3.5 (Generalization rule)**.** *If* $\vdash A$*, then* $\vdash \forall x A$*.*

Definition 3.6. The theory obtained from T by adding all of the formulas in a set of formulas Γ as non-logical axioms is called $T[\Gamma]$.

Definition 3.7 (Inconsistent)**.** A theory T is called *inconsistent* if every formula in $L(T)$ is a theorem in T. If this is not the case, T is called *consistent*.

3.2 Theorems

Theorem 3.8 (Modus Ponens)**.** *If* $\vdash A$ *and* $\vdash A \to B$*, then* $\vdash B$

Theorem 3.9 (Tautology theorem)**.** *If* B *is a tautological consequence of* $A_1, ..A_n$*, and* $\vdash A_1, ..., \vdash A_n$*, then* $\vdash B$*.*

Corollary 3.10. *Every tautology is a theorem.*

Theorem 3.11 (Substitution Theorem)**.**

1. $\vdash A[x_1 := a_1,\ ...,\ x_n := a_n] \to \exists x_1\ ...\ \exists x_n A$.
2. $\vdash \forall x_1\ ...\ \forall x_n A \to A[x_1 := a_1,\ ...,\ x_n := a_n]$.

Theorem 3.12 (Closure theorem)**.** *If* A' *is the closure of* A*, then* $\vdash A'$ *if and only if* $\vdash A$*.*

Theorem 3.13 (Deduction theorem)**.** *Let* A *be a closed formula in* $L(T)$*. For every formula* B *of* T*,* $\vdash_T A \to B$ *if and only if* B *is a theorem of* $T[A]$*.*

Theorem 3.14 (Theorem on constants)**.** *Let* T' *be obtained from* T *by adding new constants. For every formula* A *of* T *and every sequence* c_1, ..., c_n *of new constants,* $\vdash_T A$ *iff* $\vdash_{T'} A[x_1 := c_1,\ ...,\ x_n =: c_n]$

Theorem 3.15 (Equivalence theorem)**.** *Let A' be obtained from A by replacing some occurrences of B_1, ..., B_n by B'_1, ..., B'_n respectively. If $\vdash B_1 \leftrightarrow B'_1$ ... $B_n \leftrightarrow B'_n$, then $\vdash A \leftrightarrow A'$.*

Theorem 3.16 (Equality theorem)**.** *Let b' be obtained from b by replacing some of the occurrences of a_1, ..., a_n not within quantifiers with a'_1, ..., a'_n respectively, and let A' be obtained from A by the same type of replacements. If $\vdash a_1 = a'_1$, ..., $\vdash a_n = a'_n$, then $\vdash b = b'$ and $\vdash A = A'$*

Theorem 3.17 (Variant Theorem)**.** *If A' is a variant of A, then $\vdash A \leftrightarrow A'$.*

3.3 Induction on theorems

Definition 3.18 (Induction on theorems)**.** We will use *induction on theorems* to prove that every theorem in a theory T has a property P if

i) every substitution axiom, identity axiom, equality axiom and non-logical axiom has property P.

ii) if A_1, ...,A_n have property P and B is a tautological consequence of A_1, ...,A_n, then B has property P.

iii) If A has a property P and B can be inferred from A by the $\exists$-introduction rule, then B has property P

3.4 The reduction theorem

Theorem 3.19 (Reduction Theorem)**.** *Let Γ be a set of formulas in the language $L(T)$ of a theory T, and let A be a formula in $L(T)$. Then A is a theorem of $T[\Gamma]$ if and only if there is a theorem of T of the form $B_1 \to ... \to B_n \to A$, where each B_i is the closure of a formula in Γ.*

Theorem 3.20 (Reduction theorem for consistency)**.** *Let Γ be a nonempty set of formulas in the language $L(T)$ of T. Then $T[\Gamma]$ is inconsistent (3.7) if and only if there is a theorem of T which is a disjunction of negation of closures of distinct formulas in Γ.*

Corollary 3.21. *Let A' be a closure of A. Then A is a theorem of T if and only if $T[\neg A']$ is inconsistent.*

4 Consistency theorem

This chapter will prove the consistency theorem, also known as the Hilbert-Ackermann theorem. This theorem is very important for proving Herbrand's theorem, because this will give us the quasi tautology which we have to find in the 'if' side of the proof of Herbrand's theorem.

4.1 Definitions

Definition 4.1 (Extension)**.** A first order language L' is an *extension* of the first-order language L if every function, relation or constant of L is a function, relation or constant of L'.

Definition 4.2 (Conservative extension)**.** A *conservative extension* of a theory T is a theory T' such that $L(T')$ is an extension of $L(T)$ and every formula of $L(T)$ which is a theorem of T' is also a theorem of T.

Definition 4.3 (Special constants and levels)**.** Let L be a first order language. We inductively define *special constants* of level n. For $n > 0$, suppose that $\exists xA$ contains a special constant of level $n-1$, and all other special constants contained have a level of $n-1$ or lower. Then the constant $c_{\exists xA}$ is a special constant of *level* n and is called the special constant for $\exists xA$. We get the language L_c from L by adding all the special constants of all levels. The *special axiom* for $\exists xA$ will be $\exists xA \to A[x := c_{\exists xA}]$.

Definition 4.4. Let T be a theory with language L. Then T_c is a theory for L_c which we obtain by adding the special axioms as non-logical axioms.

Definition 4.5 (Belong to)**.** Let c be the special constant for $\exists xA$. Then a formula *belongs* to c if it is either the special axiom for c or a closed substitution axiom (2.20) of the form $A[x := a] \to \exists xA$

Definition 4.6 (Rank)**.** The *rank* of a special constant for $\exists xA$ is the number of occurrences of $\exists$ in $\exists xA$.

Definition 4.7. $\Delta(T)$ is the set of formulas in T_c, which are either special constants for some to some special axioms or are closed instances of identity axioms, equality axioms or non-logical axioms of T. We get $\Delta_n(T)$ from $\Delta(T)$ by only including formulas belonging to special constants with rank of at most n and the closed instances of identity axioms, equality axioms or non-logical axioms of T.

Definition 4.8 (Special Sequence)**.** We call A_1, ..., A_n a *special sequence* if $\neg A_1 \vee$ ... $\vee \neg A_n$ is a tautology (2.31).

4.2 Examples

Example 4.9. Consider the formula $\exists xA \to A[x := c_{\exists xA}]$. The special constant for this formula is $c_{\exists A \to A[x:=c_{\exists xA}]}$. Then this constant has a level of 1, since $\exists A \to A[x := c_{\exists xA}]$ contains one constant of level 0. The special constant $c_{\exists A \to A[x:=c_{\exists xA}]}$ has a rank of 1, since there is one occurrence of $\exists$ in its formula. We don't count the occurrences of $\exists$ in subscripts for ranks.

4.3 Theorems and proofs

Lemma 4.10. *Suppose that to each formula A we have associated a formula A^* so that $(\neg A)^* = \neg A^*$ and $(A \vee B)^* = A^* \vee B^*$. If B is a tautological consequence (2.29) of $A_1, .., A_n$, then B^* is a tautological consequence of A_1^*, ..., A_n^*.*

Proof. Suppose that v is a truth valuation (2.28). Define a new truth valuation v' by $v'(A) = v(A^*)$ for A elementary. We then see that $v'(A) = v(A^*)$ for all A. Thus if $v(A_1^*) = ... = v(A_n^*) = T$, then $v'(A_1) = ... = v'(A_n) = T$. So $v'(B) = T$, and thus $v(B^*) = T$ □

Lemma 4.11. *T_c (4.4) is a conservative extension of T.*

Proof. Let T' be obtained from T by adding the special constants, but not the special axioms. By the theorem on constants (3.14), T' is a conservative extension (4.2) of T. So it will suffice to show that every formula A of T which is a theorem of T_c is a theorem of T'. By the reduction theorem (3.19), we have $\vdash_{T'} B_1 \to ... \to B_n \to A$ with each B_i a distinct special axiom. We will do induction on n. For $n = 0$ we have $\vdash_{T'} A$, so we are done.

Induction step $n > 0$. Suppose we have $\vdash_{T'} B_1 \to ... \to B_n \to A$. Now suppose that the level of c, the special constant of which B_1 is the special axiom, is at least as great as the levels of the special constants for which B_2, ..., B_n are the special axioms. Then c does not occur in B_2, ..., B_n, and it does certainly not occur in A.
Then we know that B_1 is of the form $\exists x C \to C[x := c]$. So now we know that $\vdash_{T'} (\exists x C \to C[x := c]) \to ... \to B_n \to A$.

By the $\exists$-introduction rule (2.12), we know that $\vdash_{T'} \exists y(\exists x C \to C[x := y]) \to ... \to B_n \to A$.
Now, $\vdash_{T'} \exists x C \to \exists y C[x := y]$ by the variant theorem (3.17), and $\vdash_{T'} \exists y(\exists x C \to C[x := y]$ by the prenex operations (2.10.1). Now with the modus ponens (3.8), we know that $\vdash_{T'} B_2 \to ... \to B_n \to A$, and with the induction hypothesis we know that $\vdash_{T'} A$. □

Lemma 4.12. *If $\vdash_T A$ and A' is a closed instance (2.30) of A in $L(T_c)$, then A' is a tautological consequence (2.29) of formulas in $\Delta(T)$ (4.7).*

Proof. For this proof, we use induction on theorems (3.18).
If A is a substitution axiom (2.20), then A' is a closed substitution axiom (4.5) and thus in $\Delta(T)$.
If A is an identity (2.17), equality (2.18), or a non-logical axiom (2.21), then A' is a closed instance of one of those and thus in $\Delta(T)$.
If A is a tautological consequence of B_1, ..., B_n, then A' is a tautological consequence of B_1', ..., B_n', which are the closed instances of B_1, ..., B_n respectively. By the induction hypothesis, we know that each of the B_i' are tautological consequences of formulas in $\Delta(T)$, and thus so is A'.
Our final case is that A is of the form $\exists x B \to C$ and is inferred from $B \to C$ by using the $\exists$-introduction rule (2.12). Then A' is $\exists x B' \to C'$. Now, we know that $B'[x := c_{\exists x B}] \to C'$ is a closed instance of $B \to C$, and thus by the induction hypothesis a tautological consequence of formulas in $\Delta(T)$. Now A' is a tautological consequence of $\exists x B \to B'[x := c_{\exists x B}]$ and $B'[x := c_{\exists x B}] \to C'$, and

thus A' is a tautological consequence of formulas in $\Delta(T)$.
Thus with induction we have completed this proof. $\square$

Lemma 4.13. *If $n > 0$, and there is a special sequence (4.8) consisting of formulas in $\Delta_n(T)$, then there is a special sequence consisting of formulas in $\Delta_{n-1}(T)$.*

Intuition To get a special sequence in $\Delta_{n-1}(T)$ from a special sequence in $\Delta_n(T)$, we have to remove all the formulas which are in $\Delta_n(T)-\Delta_{n-1}(T)$. What are those formulas? These are the formulas which belong to special constants with a rank of exactly n. We somehow have to get rid of these formulas. We do this by removing them one by one. But, if there are more formulas belonging to a constant of rank n, we have to decide on a order in which we are going to remove said formulas. We will do this by removing the formulas belonging to the constant with the highest level first. This is so we know that these formulas can not be contained in some of the other formulas.

Proof. Suppose that there is a special sequence (4.8) consisting of formulas in $\Delta_n(T)$ (4.7). Then this special sequence consists of formulas in $\Delta_{n-1}(T)$ and formulas which belong to a special constant (4.3) of rank n (4.6). We will do this proof with induction on the number of formulas which belong to (4.5) a special constant of rank n.
If there are none, then all the formulas of the special sequence are in $\Delta_{n-1}(T)$, so there is nothing to prove.
If there is at least one, then suppose that c is a special constant with level m and c_1, ..., c_k are the other constants, that have a level which is lower than or equal to the level of c. We shall now create a special sequence consisting of formulas in $\Delta_{n-1}(T)$ and the formulas belonging to the special constants c_1, ..., c_k.

Let now A_1, ..., A_r be the formulas in the special sequence which are either in $\Delta_{n-1}(T)$ or belong to the special constants c_1, ..., c_k.
Let the remaining formulas be the special axiom
$\exists xB \to B[x := c]$ and $B[x := a_i] \to \exists xB; 1 \leq i \leq p$

We will show that there is no occurrence of $\exists xB$ in all the A_i. This is trivial if A_i is an instance (2.30) of an identity, equality or non-logical axiom, for then A_i is open. Suppose that A_i is of the form $\exists yC \to C[y := s]$ or $C[y := a] \to \exists yC$, belonging to the special constant s. Since s has a level which is equal to or lower than the level of c, $\exists xB$ has as many occurrences of $\exists$ as $\exists yC$. So $\exists xB$ cannot exist in either formula.
We now make a new special sequence where we change $\exists xB$ to $B[x := c]$ in all the formulas in our original special sequence. As proven in the above section, this does not effect the A_i. So now $\exists xB \to B[x := c]$ is changed to $B[x := c] \to B[x := c]$, which is a tautology and $B[x := a_i] \to \exists xB$ will be changed to $B[x := a_i] \to B[x := c]$. We know that this is a special sequence with lemma 4.10. So now

$$A_1, .., A_n, B[x := a_1] \to B[x := c],\ ...,\ B[x := a_p] \to B[x := c] \qquad (1)$$

We will now make a new special sequence where we will change c into a_i for each occurrence of c for a formula u, also in the subscripts of the special constants. We note this as $u^{(i)}$ Since c does not appear in B, $B[x := c]^{(i)} = B[x := a_i]$. We then obtain the special sequence

$$A_1^{(i)}, .., A_n^{(i)}, B[x := a_1]^{(i)} \to B[x := a_i],\ ...,\ B[x := a_p]^{(i)} \to B[x := a_i] \quad (2)$$

We will now claim that the sequence consisting of all the A_i and all the $A_i^{(j)}$ is a special sequence. Suppose that for a truth evaluation v (2.28), for which $v(A_i) = \mathbf{T}$ and $v(A_i^{(j)}) = \mathbf{T}$ for all i, j. Then $v(B[x := a_j]^{(i)} \to B[x := a_i]) = \mathbf{F}$ for one of these formula, since at least one of the formulas in 2 should have False assigned as truth value. So $v(B[x := a_i]) = \mathbf{F}$ for all i, so $B[x = a_i] \to B[x = c]$ will be true for all i. But then 1 is not a special sequence anymore.
What is left of the proof is showing that each $A_j^{(i)}$ is either in $\Delta_{n-1}(T)$ or is the special axiom to one of the c_1, ..., c_k. If A_j is an instance of a identity, equality or non-logical axiom, then so is $A_j^{(i)}$. Now suppose that A_j is a sentence $\exists yC \to C[y := s]$ or $C[y := a] \to \exists yC$ with s as special constant. Then $A_j^{(i)}$ is either $\exists yC^{(i)} \to C^{(i)}[y := s^{(i)}]$ or $C^{(i)}[y := a^{(i)}] \to \exists yC^{(i)}$. Since s is not c, it is clear that $s^{(i)}$ is the special constant for $\exists yC^{(i)}$. It is clear that s and $s^{(i)}$ have the same rank. So if A_j is in $\Delta_{n-1}(T)$, then so is $A_j^{(i)}$. Now suppose that s is one of the c_1, ..., c_k. Then the level of c is greater than or equal to s, so c can not appear in $\exists yC$ and thus not in s. So $s^{(i)} = s$. So $A^{(i)}$ has one of the c_1, ..., c_n as special constant.

□

Theorem 4.14 (Consistency theorem (Hilbert-Ackermann)). *An open theory (2.23) T is inconsistent if and only if there is a quasi-tautology which is a disjunction of negations of instances of non-logical axioms of T.*

Proof.
$\Longleftarrow$
Suppose that $\neg A_1 \vee ... \vee \neg A_n$ is a quasi-tautology (2.32) of which the A_i are instances of non-logical axioms of T. Then the A_i are theorems of T and so is $\neg A_1 \vee ... \vee \neg A_n$. So by the tautology theorem (3.9), T is inconsistent.

$\Longrightarrow$
Suppose T is inconsistent. Now let c be a special constant (4.3) in T_c. Since T is inconsistent, $x \neq x$ is a theorem of T. Then $c \neq c$ is an instance of $x \neq x$ and by lemma 4.12 there are formulas A_1, ..., A_n in $\Delta(T)$ (4.7) such that $A_1 \longrightarrow ... \longrightarrow A_n \longrightarrow c \neq c$. Since $c = c$ is an instance of an equality axiom, it is in $\Delta(T)$ so we can assume that one of the A_i is equal to $c = c$. Then $\neg A_1 \vee ... \vee \neg A_n$ is a tautology. So A_1, ..., A_n is a special sequence. Now there is a rank n (4.6) such that all A_i are in $\Delta_n(T)$. If we now use lemma 4.13 repeatedly, we know that there is a special sequence (4.8) in $\Delta_0(T)$, so there is a special sequence existing of identity axioms, equality axioms and non-logical axioms, and then we have a quasi tautology which is a disjunction of negations of instances of non-logical axioms of T. □

5 Herbrand's proof

This chapter will show the proof of Herbrand's theorem with the lemmas which are the most important for this proof.

Lemma 5.1. *Let T be a theory with no non-logical axioms. A closed existential (2.11) formula A is a theorem of T if and only if there is a quasi-tautology (2.32) which is a disjunction of instances (2.30) of the matrix (2.10) of A.*

Proof. Suppose that A is $\exists x_1$, ..., $x_n B$ with B open. By the corollary of the reduction theorem (3.21), A is a theorem if and only if $T[\neg A]$ is inconsistent. $T[\neg A]$ and $T[\neg B]$ are equivalent by the prenex operations and the closure theorem (3.12. So A is a theorem if and only if $T[\neg B]$ is inconsistent. With the consistency theorem (4.14) we know that $T[\neg B]$ is inconsistent if and only if there is a quasi tautology which is a disjunction of negation of closures of distinct instances of $\neg B$. So this is the case if $\neg\neg B_1$, ..., $\neg\neg B_n$ is a quasi tautology where each B_i is an instance of B. This only holds iff B_1, ..., B_n is a quasi-tautology, which completes the proof. □

Definition 5.2. We introduce an extension T'_c of T_c by adding special equality axioms which are $\forall x(A \leftrightarrow B) \rightarrow c_{\exists xA} = c_{\exists xB}$ if $c_{\exists xA}$ and $c_{\exists xB}$ are the special constants for $\exists xA$ and $\exists xB$.

Lemma 5.3. *T'_c is a conservative extension of T.*

Intuition: The idea of this proof is to take a formula A of T which is a theorem of T'_c, and we will show that this is also a theorem of T. We will do this by ordering all the special constants we have in T'_c on level. We shall show that we can make a proof of A without the special constant of the highest level, and thus we have to remove all the formulas belonging to this special constant and all the special equality axioms containing this constant. By induction, it then follows that we can rewrite the proof in T'_c to a proof in T.

Proof. We let $T[c_1, ..., c_n]$ be the theory obtained from T by adding the constants c_1, ..., c_n and the special and special equality axioms which only contain these constants. As in the proof of lemma 4.11, this will be reduced to proving the following: if $level(c_i) \leq level(c)$ for $1 \leq i \leq n$, then $T[c_1, ..., c_n, c]$ is a conservative extension of $T[c_1, ..., c_n]$.
We will do this by showing that if a formula A of $T[c_1, ..., c_n]$ has a proof in $T[c_1, ..., c_n, c]$ then it has a proof which uses no special equality axioms containing c.

We note that for each special constant appearing on the left hand side of the implication arrow of a special equality axiom $\forall x(A \leftrightarrow B) \rightarrow c_{\exists xA} = c_{\exists xB}$, the level will be lower than the constants on the right hand side of the special equality axiom. Thus in any given proof of A, we know that c can only appear on the right hand side of the equation. Now we can assume that all the formulas in the proof in which c appears are of the form $\forall x(B \leftrightarrow C) \rightarrow c = c_{\exists xC}$. It is easy to see that if a formula is of the form $\forall x(C \leftrightarrow B) \rightarrow c_{\exists xC} = c$, it can be rewritten to one of the former form. We also assume that $c = c$ does not appear on the right hand side of the equation, since that could also be derived with the equality axioms.

Let $\forall x(B \leftrightarrow C) \rightarrow c = c_{\exists xC}$ be one of the special equality axioms in the given proof of A. Let $\hat{T}$ be the theory obtained from $T[c_1, ..., c_n]$ obtained by adding the constant c and the two axioms $c = c_{\exists xC}$ and $\forall x(B \leftrightarrow C)$. We will show that A is a theorem of $\hat{T}$. For this, it will suffice to prove all the non-logical axioms in $\hat{T}$ in the given proof of A which contain c.

Since $\forall x(B \leftrightarrow C)$ is an axiom we know that $\vdash_{\hat{T}} B \leftrightarrow C$ and thus $B[x := c] \leftrightarrow C[x := c]$. Since we know that $c = c_{\exists xC}$, we also know with the equality theorem (3.16) that $\vdash_{\hat{T}} B[x := c] \leftrightarrow C[x := c_{\exists xC}]$. Since $c \neq c_{\exists xC}$, we know that $\vdash_{\hat{T}} \exists xC \rightarrow C[x := c_{\exists xC}]$, and thus with the equivalence theorem (3.15) and the fact that $\vdash_{\hat{T}} B \leftrightarrow C$ and $\vdash_{\hat{T}} B[x := c] \leftrightarrow C[x := c_{\exists xC}]$, we know that $\vdash_{\hat{T}} \exists xB \rightarrow B[x := c]$. Thus the special axiom of c is provable in $\hat{T}$.

Now consider a special equality axiom $\forall x(B \leftrightarrow D) \rightarrow c = c_{\exists xD}$ occurring in the proof of A. We know that $\forall x(C \leftrightarrow D) \rightarrow c_{\exists xC} = c_{\exists xD}$ and also because of the level of c, we know that c does not occur in this formula. Now since $\vdash_{\hat{T}} B \leftrightarrow C$, we know with the equivalence theorem (3.15) that $\vdash_{\hat{T}} \forall x(B \leftrightarrow D) \rightarrow c_{\exists xC} = c_{\exists xD}$. Because $\vdash_{\hat{T}} c = c_{\exists xC}$, we also know that $\vdash_{\hat{T}} \forall x(B \leftrightarrow D) \rightarrow c = c_{\exists xD}$.

Since $\vdash_{\hat{T}} A$, we know with the deduction theorem (3.13) that $\vdash_T c = c_{\exists xC} \rightarrow \forall x(B \leftrightarrow C) \rightarrow A$ and with the theorem on constants (3.14) that $\vdash_{T[c_1,...,c_n]} y = c_{\exists xC} \rightarrow \forall x(B \leftrightarrow C) \rightarrow A$. Now we substitute $c_{\exists xC}$ for y and since $\vdash c_{\exists xC} = c_{\exists xC}$ with the identity axiom we know that $\vdash_{T[c_1,...,c_n]} \forall x(B \leftrightarrow C) \rightarrow A$.
Now by using $\neg(\forall x(B \leftrightarrow C) \rightarrow c = c_{\exists xC}) \rightarrow \forall x(B \leftrightarrow C)$ by example 2.27 and the tautology theorem (3.9), we know that
$\vdash_{T[c_1,...,c_n,c]} \neg(\forall x(B \leftrightarrow C) \rightarrow c = c_{\exists xC}) \rightarrow A$ without using any nonlogical axioms containing c.
Now let D_1, ..., D_k be the special equality axioms containing c which are used in the given proof of A. By the deduction theorem (3.13) $D_1 \rightarrow ... \rightarrow D_k \rightarrow A$ has a proof not using special equality axioms containing c. In the above we have shown that each $\neg D_i \rightarrow A$ also has a proof without c. Thus by the tautology theorem (3.9) A has a proof without c. ☐

Remark 5.3.1. Let $c_{\exists x \neg A}$ be a constant for $\exists x \neg A$.
Then $\exists x \neg A \rightarrow \neg A[x := c_{\exists x \neg A}]$ is an axiom of T'_c.
Thus we know that $\vdash_{T'_c} \exists x \neg A \rightarrow \neg A[x := c_{\exists x \neg A}]$, and using $a \rightarrow b \leftrightarrow \neg b \rightarrow \neg a$ (proof by contradiction), we know that $\vdash_{T'_c} \neg\neg A[x := c_{\exists x \neg A}] \rightarrow \neg \exists x \neg A$.
Now since we know that $\neg\neg A \rightarrow A$ (by example (2.26) and $\neg \exists x \neg A \leftrightarrow \forall A$ (by subsection (2.2)), we know that $\vdash_{T'_c} A[x := c_{\exists x \neg A}] \rightarrow \forall x A$.

Definition 5.4 (A^* and A_H)**.** Let A be be a closed formula in prenex form. If A is existential, then A_H is equal to A.
If not, A is of the form $\exists x_1 ... \exists x_n \forall y B[x_1, ..., x_n, y]$. Then let f be a new n-ary function and be A^* be $\exists x_1 ... \exists x_n B[x_1, ..., x_n, y := f(x_1, ..., x_n)]$. Now if A^* is existential, A_H will be A^*. Otherwise, repeat this process of making A^{**}, A^{***}, etc until the result is existential. The result will be A_H.

Example 5.5. Let A be $\exists x \forall y \exists z \forall w B[x, y, z, w]$. Then A^* will be $\exists x \exists z \forall w B[x, f(x), z, w]$ and A^{**} will be $\exists x \exists z B[x, f(x), z, g(x, z)]$. This last formula is existential so A_H will be equal to A^{**}. The matrix of A_h will be $B[x, f(x), z, g(x, z)]$.

Theorem 5.6 (Herbrand's Theorem)**.** *Let T be a theory with no non-logical axioms, and let A be a closed formula in prenex form in T. Then A is a theorem of T if and only if there is a quasi-tautology which is a disjunction of instances of the matrix of A_H.*

Intuition: This proof is done as follows: The 'if' side of the proof follows almost directly, so we do this side first.
The interesting part of this proof is the 'and only if' side, this is the process which is called deherbrandisation. We first show that A can be proven from an instance of the matrix of A, in which we use some terms $a_1, a_2, ..., a_n$ and some constants $c_1, c_2, ..., c_n$ We assume that we have a quasi tautology which is a disjunction of instances of the matrix of A_H. Then we know that this is a tautological consequence of instances of identity and equality axioms, which we will call $C_1, ..., C_r$. We will now substitute in this quasi tautology all instances of the functions $f_1, ..., f_n$ gained by Herbrandisation applied on the terms $a_1, ..., a_n$ by the constants $c_1, ..., c_n$, until all appearances of the f_i are gone, also where they appear in the a_i. We will now repeat this process on the axioms C_i, thus gaining the equality and identity axioms of which we found a tautological consequence for the substituted quasi-tautology, which is thus also a quasi-tautology. This will complete the proof.

Proof.
$\Longleftarrow$
Suppose there is a quasi-tautology which is a disjunction of instances of the matrix of A_H. Let T' be obtained by adding the new function symbols of A_H. Suppose A is of the form $\forall y B$, then we know that $\vdash_{T'} \forall y B \longrightarrow B[y := f(x_1, ..., x_n)]$ by the substitution theory (3.11). Since $A^* := B[y := f(x_1, ..., x_n)]$, we know that $\vdash_{T'} A \longrightarrow A^*$. By repeating this process, we can deduce that $\vdash_{T'} A \longrightarrow A_H$. If $\vdash_T A$, then $\vdash_{T'} A$ because T' is a conservative extension, so then by modus ponens (3.8) we know that $\vdash_{T'} A_H$.

$\Longrightarrow$
Suppose, to simplify the description, that A is of the form $\exists x \forall y \exists z \forall w B[x, y, z, w]$ with B open. Then A_H will be $\exists x \exists z B[x, f(x), z, g(x, z)]$. T_c is a conservative extension of T in the non-logical axioms are the non-logical axioms of T and the special axioms $\exists x A \to A[x := c_{\exists x A}]$ for the special constants of L (4.3). Then let a and b be variable free terms in T'_c, and define c_a and $c_{a,b}$ as the special constants respectively for $\exists y \neg \exists z \forall w B[a, y, z, w]$ and $\exists w \neg B[a, c_a, b, w]$.

Then we know by the remark 5.3.1 and the substitution theorem (3.11) that
$\vdash_{T'_c} B[a, c_a, b, c_{a,b}] \to \forall w B[a, c_a, b, w]$
$\vdash_{T'_c} \forall w B[a, c_a, b, w] \to \exists z \forall w B[a, c_a, z, w]$
$\vdash_{T'_c} \exists z \forall w B[a, c_a, b, w] \to \forall y \exists z \forall w B[a, y, z, w]$
$\vdash_{T'_c} \forall x \exists z \forall w B[a, y, z, w] \to \exists x \forall y \exists z \forall w B[x, y, z, w]$

And thus $\vdash B[a, c_a, b, c_{a,b}] \to A$. Using the special equality axiom from T'_c which is $\forall x(A \leftrightarrow B) \to c_{\exists x A} = c_{\exists x B}$, we know that if $a = a', b = b'$ then
$\forall y(\neg \exists z \forall w B[a, y, z, w] \leftrightarrow \neg \exists z \forall w B[a', y, z, w]) \to c_a = c_{a'}$ and
$\forall w(\neg B[a, c_a, b, w] \leftrightarrow \neg B[a', c_{a'}, b', w]) \to c_{a,b} = c_{a',b'}$.
And thus $\vdash a = a' \to c_a = c_{a'}$ and $\vdash a = a' \to b = b' \to c_{a,b} = c_{a',b'}$.

Now assume that $\vdash_{T'} A_H$. By lemma 5.1, there is a quasi tautology which is a disjunction of instances of the matrix:

$$B[a_1, f(a_1), b_1, g(a_1, b_1)] \vee ... \vee B[a_n, f(a_n), b_n, g(a_n, b_n)] \tag{3}$$

in which a_i, b_i are terms of $L(T')$. Next we substitute all $f(a)$ by c_a and all $g(a,b)$ by $c_{a,b}$, where f and g do not occur in a and b. We do this until there are no occurrences of f and g. This will result in the following formula:

$$B[a'_1, c_{a'_1}, b'_1, c_{a'_1,b'_1}] \vee ... \vee B[a'_n, c_{a'_n}, b'_n, c_{a'_n,b'_n}] \tag{4}$$

We will now show that this is a theorem of T'_c.

The formula 3 is a tautological consequence (2.29) of instances C_1, ..., C_r of identity and equality axioms. Now, we can substitute $f(a)$ and $g(a,b)$ by c_a and $c_{a,b}$ like we did when transforming formula 3 to 4. This will result into formulas C'_1 ,..., C'_r. By a lemma 4.10 it holds that 4 is a tautological consequence of the C'_i. The C'_i will either be an equality or identity axiom, or from the form $x = x' \rightarrow f(x) = f(x')$ or $x = x' \rightarrow y = y' \rightarrow g(x,y) = g(x',y')$. But we have proven that these are theorems. Then it will follow from $\vdash B[a, c_a, b, c_{a,b}] \rightarrow A$ and the tautology theorem that A is a theorem of T'_c, and by lemma 5.3 it will also be a theorem of T. $\square$

Example 5.7 (Proving the drinker's paradox)**.** We will now view the example of the drinker's paradox. Suppose the formula $A := \exists x(D(x) \to \forall y D(y))$ in which $D(x)$ is the relation 'Person x is drinking'. Then this formula means 'There is a person x in the pub, such that, when he or she is drinking, everybody in the pub is drinking'. This seems counter intuitive, but consider the following cases: Suppose that everybody in the pub is drinking. Then you can point at a random person who is drinking, and he or she will be the person such that when he or she is drinking, everybody is drinking. So the theorem will hold in this case
Now suppose that not everybody is drinking. Then you can point at the person who is not drinking. Since false $\to$ false $=$ true, the theorem will still hold.

We will now prove the drinker's paradox using Herbrand's theorem. So again, our theorem is $\exists x(D(x) \to \forall y D(y))$. First of all, we want to write this in prenex form. This will give us the following formula: $\exists x \forall y(D(x) \to D(y))$. We will now rewrite the part after the quantifiers using the rules in 2.2). This will give us $\exists x \forall y(\neg D(x) \vee D(y))$.

Now, we will use Herbrand's theorem to replace y by introducing a function f of x, this will give us $\exists x(\neg D(x) \vee D(f(x))$. Now, using Herbrand's theorem, we know that we can prove this by proving a disjunction of instances of the matrix, which is $(\neg D(x) \vee D(f(x))$. We will choose to substitute x once by x and once by $f(x)$, so this will give us $(\neg D(x) \vee D(f(x))) \vee (\neg D(f(x)) \vee D(f(f(x))))$. Now rearranging the brackets will give us $\neg D(x) \vee (D(f(x)) \vee \neg D(f(x))) \vee D(f(f(x)))$, which is obviously true.

Example 5.8 (The drinker's paradox: deherbrandisation)**.** Now, we will show an example of the other side of the proof. Suppose that we have a proof for

$$(\neg D(x) \vee D(f(x))) \vee (\neg D(f(x)) \vee D(f(f(x))))$$

then we want to get a proof for

$$\exists x(D(x) \to \forall y D(y))$$

without using $f(x)$.

Let c_a be the special constant belonging to $\exists y \neg(\neg D(a) \vee D(y))$ and let c_{c_a} be the special constant belonging to $\exists y \neg(\neg D(c_a) \vee D(y))$.
Assume that $\neg D(x) \vee D(f(x))$ holds.
We know that

$$\exists y \neg(\neg D(a) \vee D(y)) \to \neg(\neg D(a) \vee D(c_a))$$

because this is the special axiom for c_a. By the fact that

$$(A \to B) \leftrightarrow (\neg B \to \neg A)$$

we know that

$$\neg\neg(\neg D(a) \vee D(c_a)) \to \neg \exists y \neg(\neg D(a) \vee D(y))$$

Now by the operations in subsection 2.2 and the fact that $\neg\neg A \leftrightarrow A$ we know that

$$(\neg D(a) \vee D(c_a)) \to \forall y(\neg D(a) \vee D(y))$$

Now by the substitution theorem (3.11) we know that

$$\forall y(\neg D(a) \vee D(y)) \rightarrow \exists x \forall y(\neg D(x) \vee D(y))$$

We also know that

$$\exists x \forall y(\neg D(x) \vee D(y)) \rightarrow \exists x(D(x) \rightarrow \forall y D(y)))$$

by subsection 2.2.
Now we know that

$$(\neg D(a) \vee D(c_a)) \rightarrow \exists x(D(x) \rightarrow \forall y D(y))$$

Using the same method we also know that

$$(\neg D(c_a) \vee D(c_{c_a})) \rightarrow \exists x(D(x) \rightarrow \forall y D(y))$$

Now we have a proof for

$$(\neg D(x) \vee D(f(x)) \vee (\neg D(f(x)) \vee D(f(f(x))))$$

We can make a proof for

$$(\neg D(a) \vee D(c_a)) \vee (\neg D(c_a) \vee D(c_{c_a}))$$

in T_c' from this by substituting x by a, $f(x)$ by c_a and $f(f(x))$ by c_{c_a} in the entire proof. Now we know that T_c' is a conservative extension of T, so there also exists a proof in T. So we know that if we have a proof for

$$(\neg D(x) \vee D(f(x)) \vee (\neg D(f(x)) \vee D(f(f(x))))$$

we can transform this into a proof of $\exists x(D(x) \rightarrow \forall y D(y))$.

6 Conclusion and further reading

In this thesis, we have given a proof of Herbrand's theorem. With this, we know that is is possible to deherbrandise a formula, when only equality and identity axioms are used.

In this thesis, only the mathematical side of Herbrand's theorem is discussed. If the reader is interested in an actual implementation of deherbrandisation, one can look at the bachelor thesis of Ramon van Sparrentak from 2014. He discusses an implementation of deskolemisation, the dual of deherbrandisation.

If the reader wants to read more information on the proofs which were not done in this thesis or further reading on this topic, then the reader should look up Mathematical Logic by Joseph R. Shoenfield. This is the book which has been primarily used for this thesis.

References

[1] Shoenfield, J.R (1967) *Mathematical Logic*. Massachusetts, Addison-Wesly Publishing Company

[2] Weller, D (2012) Deskolemization, Equality and Logical Complexity [Presentation]. Retrieved from `https://www.logic.at/staff/weller/lmrc2012.pdf`

[3] Baaz, M., Hetzl, S., & Weller, D. (2012). On the complexity of proof deskolemization. *The Journal of Symbolic Logic*, 77(02), 669-686

[4] Sparrentak, R (2014) A concrete deskolemization algorithm [Bachelor Thesis]. Retrieved from `http://www.cs.ru.nl/bachelorscripties/#2014`