Between 1902 and 1908 Bertrand Russell proposed various “theories of type” in response to his discovery that Gottlob Frege’s version of naive set theory was afflicted with Russell’s paradox. By 1908 Russell arrived at a “ramified” theory of types together with an “axiom of reducibility” both of which featured prominently in Whitehead and Russell’s Principia Mathematica published between 1910 and 1913. They attempted to resolve Russell’s paradox by first creating a hierarchy of types, then assigning each concrete mathematical (and possibly other) entity to a type. Entities of a given type are built exclusively from entities of those types that are lower in their hierarchy, thus preventing an entity from being assigned to itself.[1]

With Russell’s discovery (1901, 1902)[2] of a paradox in Gottlob Frege’s 1879 Begriffsschrift and Frege’s acknowledgment of the same (1902), Russell tentatively introduced his solution as “Appendix B: Doctrine of Types” in his 1903 The Principles of Mathematics.[3] This contradiction can be stated as “the class of all classes that do not contain themselves as elements”.[4] At the end of this appendix Russell asserts that his “doctrine” would solve the immediate problem posed by Frege, but “there is at least one closely analogous contradiction which is probably not soluble by this doctrine. The totality of all logical objects, or of all propositions, involves, it would seem a fundamental logical difficulty. What the complete solution of the difficulty may be, I have not succeeded in discovering; but as it affects the very foundations of reasoning...”

By the time of his 1908 Mathematical logic as based on the theory of types[6] Russell had studied “the contradictions” (among them the Epimenides paradox, the Burali-Forti paradox, and Richard’s paradox) and concluded that “In all the contradictions there is a common characteristic, which we may describe as self-reference or reflexiveness”.

In 1903, Russell defined predicative functions as those whose order is one more than the highest-order function occurring in the expression of the function. While these were fine for the situation, impredicative functions had to be disallowed:

A function whose argument is an individual and whose value is always a first-order proposition will be called a first-order function. A function involving a first-order function or proposition as apparent variable will be called a second-order function, and so on. A function of one variable which is of the order next above that of its argument will be called a predicative function; the same name will be given to a function of several variables [etc].

He repeats this definition in a slightly different way later in the paper (together with a subtle prohibition that they would express more clearly in 1913):
A predicative function of x is one whose values are propositions of the type next above that of x, if x is an individual or a proposition, or that of values of x if x is a function. It may be described as one in which the apparent variables, if any, are all of the same type as x or of lower type; and a variable is of lower type than x if it can significantly occur as argument to x, or as argument to an argument to x, and so forth.

This usage carries over to Alfred North Whitehead and Russell’s 1913 Principia Mathematica wherein the authors devote an entire subsection of their Chapter II: “The Theory of Logical Types” to subchapter I.[2]
Russell’s paradox is the most famous of the logical or set-theoretical paradoxes. Also known as the Russell-Zermelo paradox, the paradox arises within naïve set theory by considering the set of all sets that are not members of themselves. Such a set appears to be a member of itself if and only if it is not a member of itself. Hence the paradox.

---

[3]:

Central to any theory of sets is a statement of the conditions under which sets are formed. In addition to simply listing the members of a set, it was initially assumed that any well-defined condition (or precisely specified property) could be used to determine a set.

More precisely, naïve set theory assumes the so-called naïve or unrestricted Comprehension Axiom.

But from the assumption of this axiom, Russell's contradiction follows.

As Russell tells us, it was after he applied the same kind of reasoning found in Cantor's diagonal argument to a "supposed class of all imaginable objects" that he was led to the contradiction:

"The comprehensive class we are considering, which is to embrace everything, must embrace itself as one of its members. In other words, if there is such a thing as "everything," then, "everything" is something, and is a member of the class "everything." But normally a class is not a member of itself. Mankind, for example, is not a man. Form now the assemblage of all classes which are not members of themselves. This is a class: is it a member of itself or not? If it is, it is one of those classes that are not members of themselves, i.e., it is not a member of itself. If it is not, it is not one of those classes that are not members of themselves, i.e. it is a member of itself. Thus of the two hypotheses – that it is, and that it is not, a member of itself – each implies its contradictory. This is a contradiction. (1919, 136)"

Standard responses to the paradox attempt to limit in some way the conditions under which sets are formed. The goal is usually both to eliminate R (and similar contradictory sets) and, at the same time, to retain all other sets needed for mathematics. This is often done by replacing the unrestricted Comprehension Axiom with the more restrictive Separation Axiom.

Unlike Burali-Forti's paradox, Russell's paradox does not involve either ordinals or cardinals, relying instead only on the primitive notions of set and set inclusion.

The significance of Russell's paradox can be seen once it is realized that, using classical logic, all sentences follow from a contradiction.

Because set theory underlies all branches of mathematics, many people began to worry that the inconsistency of set theory would mean that no mathematical proof could be completely trustworthy. Only by eliminating Russell's paradox could mathematics as a whole regain its consistency.

Russell's paradox ultimately stems from the idea that any condition or property may be used to determine a set.

Russell's own response to the paradox came with his aptly named theory of types. Believing that self-application lay at the heart of the paradox, Russell's basic idea was that we can avoid commitment to R (the set of all sets that are not members of themselves) by arranging all sentences (or, more precisely, all propositional functions, functions which give propositions as their values) into a hierarchy. It is then possible to refer to all objects for which a given condition (or predicate) holds only if they are all at the same level or of the same "type."

This solution to Russell's paradox is motivated in large part by adoption of the so-called vicious circle principle. The principle in effect states that no propositional function can be defined prior to specifying the function's scope of application. In other words, before a function can be defined, one must first specify exactly those objects to which the function will apply (the function's domain). For example, before defining the predicate "is a prime number," one first needs to define the collection of objects that might possibly satisfy this predicate, namely the set, N, of natural numbers.

As Whitehead and Russell explain, "An analysis of the paradoxes to be avoided shows that they all result from a kind of vicious circle. The vicious circles in question arise from supposing that a collection of objects may contain members which can only be defined by means of the collection as a whole. Thus, for example, the collection of propositions will be supposed to contain a proposition stating that "all propositions are either true or false." It would seem, however, that such a statement could not be legitimate unless "all propositions" referred to some already definite collection, which it cannot do if new propositions are created by statements about "all propositions." We shall, therefore, have to say that statements about "all propositions" are meaningless. … The principle which enables us to avoid illegitimate totalities may be stated as follows: “Whatever involves all of a collection must not be one of the collection”; or, conversely: “If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total.” We shall call this the “vicious-circle principle,” because it enables us to avoid the vicious circles involved in the assumption of illegitimate totalities.” (1910, 2nd edn 37)

If Whitehead and Russell are right, it follows that no function’s scope of application will ever be able to include any object presupposed by the function itself. As a result, propositional functions (along with their corresponding propositions) will end up being arranged in a hierarchy of the kind Russell proposes.

Although Russell first introduced his theory of types in his 1903 Principles of Mathematics, he recognized immediately that more work needed to be done since his initial account seemed to resolve some but not all of the paradoxes. Among the alternatives he considered was a so-called substitutional theory (Galaugher 2013). This in turn led to type theory’s more mature expression five years later in Russell’s 1908 article, “Mathematical Logic as Based on the Theory of Types,” and in the monumental work he co-authored with Alfred North Whitehead, Principia Mathematica (1910, 1912, 1913). Russell’s type theory thus appears in two versions: the “simple theory” of 1903 and the “ramified theory” of 1908. Both versions have been criticized for being too ad hoc to eliminate the paradox successfully.

Russell’s paradox is sometimes seen as a negative development – as bringing down Frege’s Grundgesetze and as one of the original conceptual sins leading to our expulsion from Cantor’s paradise. (But) for one thing, although the matter remains controversial, later research has revealed that the paradox does not necessarily short circuit Frege’s derivation of arithmetic from logic alone. Frege’s version of NC (his Axiom V) can simply be abandoned. (For details, see the entry on Frege’s Theorem.) For another, Church gives an elegant formulation of the simple theory of types that has proven fruitful even in areas removed from the foundations of mathematics. (For details, see the entry on Type Theory.) Finally, the development of axiomatic (as opposed to naïve) set theories which exhibit various ingenious and mathematically and philosophically significant ways of dealing with Russell’s paradox paved the way for stunning results in the metamathematics of set theory. These results have included Gödel’s and Cohen’s theorems on the independence of the axiom of choice and Cantor’s continuum hypothesis.

Can someone recommend me an English-language version of the Begriffsschrift in modern notation.
It’d probably be best just to learn how to read it. Roy Cook has a very thorough appendix to the recent translation of the Grundgesetze which will teach you everything you need to know.

1 https://en.wikipedia.org/wiki/Type_theory

2 https://en.wikipedia.org/wiki/Axiom_of_reducibility

3 Russell's Paradox (Stanford Encyclopedia of Philosophy_Winter 2014 Edition)