Replizierbarkeit und Reproduzierbarkeit

Bevor wir uns anschauen, wie wir Daten erheben und mit ihnen arbeiten können, wollen wir uns kurz mit zwei wichtigen Prinzipien beschäftigen, die bei der Arbeit mit Daten von zentraler Bedeutung sind: **Replizierbarkeit** und **Reproduzierbarkeit**.



[Bild: Replizierbarkeit und Reproduzierbarkeit/meerkat-255564\_1920]

Was ist damit gemeint? Beiden Prinzipien liegt ein gemeinsames Ziel zugrunde:



[Bild: Replizierbarkeit und Reproduzierbarkeit/raquel-martinez-SQM0sS0htzw-unsplash]

Die **Nachvollziehbarkeit** und Nachprüfbarkeit unserer Datenarbeit muss gesichert sein. Dieses Grundprinzip kennen Sie auch aus anderen Bereichen des wissenschaftlichen Arbeitens: Wenn Sie in einer Hausarbeit eine Aussage belegen möchten, ist so etwas wie “Ich meine mich zu erinnern, das irgendwo gelesen zu haben” keine ausreichende Literaturangabe.

Bei datenbasierter Arbeit gilt ebenso, dass Sie die Leserin bzw. den Leser in die Lage versetzen müssen, die Erhebung, die Sie vorgenommen haben, prinzipiell selbst vornehmen zu können. Daher ist es wichtig, genau zu beschreiben, wie die Daten, mit denen Sie arbeiten, gewonnen wurden und wie Sie sie analysiert haben.

[Bild: Replizierbarkeit und Reproduzierbarkeit/ashish-ansurkar-nRS8kyWaexs-unsplash]

Am besten ist es, wenn Sie die Rohdaten Ihrer Arbeit zugänglich machen. Bei Hausarbeiten können Sie z.B. den jeweiligen Dozierenden Ihre Daten per Mail schicken. Bei wissenschaftlichen Veröffentlichungen empfiehlt es sich, die Rohdaten in einem sog. Repositorium hochzuladen, etwa <a href="https://osf.io/" target="blank">OSF</a> (fachübergreifend) oder <a href="https://dataverse.no/dataverse/trolling" target="blank">TROLLing</a> (spezifisch linguistisch). Bei Korpusanalysen können die Daten, die Sie hochladen, beispielsweise Beleglisten sein, bei Fragebogenstudien die (anonymisierten) Ergebnisse Ihrer Fragebogenstudie.

Damit sichern Sie die Reproduzierbarkeit und Replizierbarkeit Ihrer Studie.



[Bild: Replizierbarkeit und Reproduzierbarkeit/bank-phrom-Tzm3Oyu\_6sk-unsplash]

Mit **Reproduzierbarkeit** ist gemeint, dass eine andere Forscherin mit denselben Daten zu den gleichen Ergebnissen kommt.



[Bild: Replizierbarkeit und Reproduzierbarkeit/laboratory-2815641\_1920]

Mit **Replizierbarkeit** ist gemeint, dass unter den gleichen Bedingungen, unter denen Sie gearbeitet haben, neue Daten erhoben werden können, um die Ergebnisse, die auf Grundlage der neuen Daten erzielt werden, mit Ihren Ergebnissen zu vergleichen.

Ein einfaches Beispiel: Julia hat eine Korpusstudie zur graphematischen Variation bei *zu Hause / zuhause / Zuhause* gemacht. Wenn ich weiß,

* welches Korpus sie verwendet hat und
* welche Suchanfrage sie gestellt hat,

kann ich die Studie eins zu eins “nachbauen”. Ich kann sie also **reproduzieren**. Wenn Julia komplexere statistische Analysen benutzt hat, muss ich natürlich zusätzlich noch wissen, wie genau sie dabei vorgegangen ist.

Wenn ich eine Studie reproduzieren kann, kann ich sie auch **replizieren**, also die gleiche Studie noch einmal durchführen, aber mit anderen Daten, zum Beispiel mit einem neuen Korpus.

[Bild: Replizierbarkeit und Reproduzierbarkeit/lightbulb-2692247\_1920]

Je nachdem, um welche Art von Daten es sich handelt, kann es natürlich sein, dass Ihre Ergebnisse nicht exakt replizierbar sind: Bei einer Fragebogenstudie beispielsweise werden sich immer wieder leichte Verschiebungen in der Zusammensetzung der Teilnehmenden ergeben. Genau darin liegt aber auch ein wichtiger Mehrwert von Replikationsstudien, denn wenn sich die Ergebnisse deutlich von denen aus vorheriger Forschung unterscheiden, kann man z.B. überprüfen, ob das an einer Veränderung in Eigenschaften der Stichprobe liegt, die zuvor nicht berücksichtigt wurden.