



Kann die Vermessung der Sprache eines Bewerbers bei der Personalauswahl helfen? Linguisten und Psychologen zweifeln dies an.

# Personalauswahl per Sprachtest

**KRITIK.** Ein Sprachtest soll angeblich Rückschlüsse auf Persönlichkeitsmerkmale ziehen können und so bei der Personalauswahl helfen. Experten halten das für fragwürdig.

Von **Bärbel Schwertfeger**

Ich tippe den Code ins Telefon, drücke die Raute-Taste. Dann starten die Fragen, die ich beantworten soll und mit deren Hilfe bis zu 220.000 Features meiner Sprache und damit auch meine Persönlichkeit erfasst werden sollen. Bitte beschreiben Sie den Ablauf eines typischen Sonntags. Was macht Ihnen in Ihrem aktuellen oder erlernten Beruf am meisten Freude? Bitte beschreiben Sie ausführlich, wie, wo und auf welche Weise Sie am besten entspannen können. Wie sollte ein idealer Mitarbeiter sein? Die Fragen wiederholen sich mehrmals in ähnlicher Form. Dreimal soll ich antworten, wie ich Weihnachten verbracht habe. Das nervt nicht nur mich. „Das Feedback bekommen wir häufiger“, erklärt Christian Greb, der beim Unternehmen Psyware für die Entwicklung des Sprachtests Pre-

precire verantwortlich ist, später im Auswertungsgespräch. Die Fragen würden eben so oft eingespielt, bis man eine Sprachprobe von 12 bis 15 Minuten habe.

## Faszinierendes Produktversprechen

Was Precire verspricht, klingt faszinierend. Über die automatisierte Erfassung von Sprache sollen sich unmittelbare Rückschlüsse „auf Persönlichkeitsmerkmale, die psychische Belastung sowie die anwendungsspezifische Eignung des Sprechenden“ ziehen lassen. Big Data macht's möglich. Also Schluss mit aufwändigen Assessment Center oder teuren Management Audits? Genügt es künftig, wenn der Kandidat 15 Minuten mit dem Computer spricht?

„Das Ergebnis beruht auf der Sprache und der Vermessung der Persönlichkeit“, erklärt Greb. Dabei werde das Sprachprofil auf Zahlenwerte herunter-

gebrochen, die dann wiederum zu Merkmalen verdichtet werden. „Wenn sich die Schuhgröße in der Sprache zeigen würde, versuchen wir sprachliche Zusammenhänge zu finden, die zu einer Schuhgröße passen“, erklärt Greb. Bisher hätten 5.000 Personen den Sprachtest und gleichzeitig einen Persönlichkeitstest gemacht. Durch die Korrelation ließe sich so anhand dieses Referenzprofils die Persönlichkeit der Testperson erkennen. Das sei ähnlich wie ein Persönlichkeitstest, erklärt Greb. Nur könne der Bewerber das Ergebnis des Sprachtests nicht manipulieren. Im Unternehmen könnten auch die erfolgreichen Mitarbeiter als Referenzgruppe dienen. „Identifizieren Sie Menschen, die zu Ihrer Unternehmenskultur passen und über die Traits und Skills verfügen, welche Sie benötigen“, heißt es dazu auf der Website der Firma Psyware, die den Precire-Test anbietet. Dahinter steckt ein

---

## Valide ist ein Test erst dann, wenn von neutralen Wissenschaftlern nachgewiesen ist, dass er auch tatsächlich das misst, was er vorgibt zu messen.

---

mit mehreren Millionen Euro von Investoren gefördertes Start-up mit Psychologen, IT-Experten, Mathematikern und Sprachwissenschaftlern. Geschäftsführer ist der Jurist Dirk Gratzel.

Zurück zum Selbsttest. Eine halbe Stunde vor dem telefonischen Auswertungsgespräch bekomme ich ein siebenseitiges Dokument mit vielen Grafiken, Zahlen und Fremdwörtern wie audio-prosodische Parameter oder deiktische Verteilung zugeschickt. Ich verwende das Wort „ich“ und „man“ häufiger als es der Referenzwert angibt. Führungskräfte, die häufig „man“ benutzen, tun sich schwer, Bindung zu erzeugen, erklärt Psychologe Greb. Bei meinem persönlichen Sprachprofil nach dem Modell von „Schulz von Thun“ bin ich wiederum – trotz häufigem „man“ – sehr beziehungsorientiert. Erstaunlich auch mein „Psycholinguistischer Resilienz-Index“. Obwohl mir zuvor ein im Vergleich zur Referenzgruppe deutlich höherer Anteil bei den Negationen bestätigt wurde, heißt es nun, dass ich „vermehrt positive Begrifflichkeiten nutze“. In der Kurzzusammenfassung bekomme ich den Entwicklungstipp: „Reflektieren Sie im Alltag, wie häufig Verneinungen vorkommen.“ Ja was denn nun? Beim Profilvergleich werden meine Sprechmerkmale mit dem Big-Five-Persönlichkeitsmodell verglichen. Ich habe eine Übereinstimmung von 71,97 Prozent mit dem Referenzprofil. Im Durchschnitt sind es 77,84 Prozent. Das liege an meiner stark ausgeprägten Autonomie, sagt Greb. Überzeugend ist das für mich nicht. Umso erstaunlicher ist es, dass Psyware-Geschäftsführer Gratzel in einem Interview behauptet, die Vorhersagen seien zu 85 bis 90 Prozent richtig.

### Experten hegen Zweifel

Jarek Krajewski, Professor am Institut für Sicherheitstechnik, Abteilung Human Factors und Diagnostik an der Universität Wuppertal, hält das „wissenschaftlich für unseriös“. „Bei der Entschlüsselung der Sprachmerkmale beißen sich die weltweit

besten Wissenschaftler noch die Zähne aus“, erklärt der Psychologe, dessen Forschungsschwerpunkte die akustische Stimmanalyse und Sprachemotionserkennung ist. „Seriös bewertet, befänden sich die Systeme derzeit bestenfalls in einem Prototyp-Stadium. Selbst die besten sprachbasierten Instrumente kämen unter realistischen, replizierbaren Bedingungen derzeit allenfalls auf 70 bis 80 Prozent Vorhersagegüte. Zudem sei Sprache sehr situationsabhängig. Die Ergebnisse hängen stark von der Sprachaufgabe und der Sprechsituation ab“, sagt der Psychologe. „Und die Aussage, dass man Sprache nicht willkürlich manipulieren kann, sei zwar ein cleveres Marketinginstrument, aber schlichtweg Unsinn.“

„Um herauszufinden, ob ein Messverfahren wirklich funktioniert, muss es von unabhängigen Wissenschaftlern überprüft werden“, erklärt Professor Krajewski. „Solange eine seriöse und unabhängige Validierung fehlt, kann man auch gleich aus der Hand lesen.“ Valide ist ein Test erst dann, wenn nachgewiesen ist, dass er auch das misst, was er vorgibt zu messen. Zwar zweifle er nicht daran, dass es signifikante Korrelationen zwischen Sprachmerkmalen und diversen Persönlichkeitsmerkmalen und psychologischen Zuständen gebe, so Krajewski, doch die seien in der Regel schwach, so dass eine Prognose mit Fehlern behaftet sei. „Das gilt auch, wenn man viele kleine, jeweils einzeln schwach korrelierende Stimmerkmale regressionsanalytisch zusammenfasst. Wenn es nur eine sehr dünne theoretische Grundlage und keine neutral überprüfte, saubere Empirie gebe, dann sei das aus wissenschaftlicher Perspektive schon ziemlich dürftig. „Das ist derzeit zu 95 Prozent

Marketing und zu 5 Prozent Substanz“, behauptet der Wissenschaftler und rät daher, „lieber die Finger davon zu lassen und einen etablierten Persönlichkeitstest einsetzen solange die Ergebnisse nicht von einer neutralen Drittinanz repliziert werden.“

### Noch keine ausreichende Validierung

Doch Psyware legt auch auf mehrmalige Nachfrage keine Validierungsstudien vor. Es liefen gerade weitere Studien, erklärt Greb. Eine unabhängige wissenschaftliche Veröffentlichung gebe es bisher nicht. Lediglich ein Studienbericht „Validierung des Precire-VoiceCheck“ der Ludwig-Maximilians-Universität München in Studienkooperation mit Psyware GmbH und Mind & Mood GmbH wird zur Verfügung gestellt, in der eine auffallend hohe Korrelation von 0,7 zwischen Precire und Stressfragebögen angegeben wird. Das sei für ihn „hochgradig unplausibel“, so Professor Krajewski. So erzielten die weltweit erfolgreichsten Gruppen in objektiven, das heißt unter realistischen, überprüfbaren Bedingungen durchgeführte Sprachsignal-Wettbewerbe lediglich Validitätskorrelationen von 0,4 bis 0,5.

Psyware beruft sich auf die sogenannte interne Validität. Dabei korreliert man die Sprachprobe mit einem etablierten Persönlichkeitstest und schließt dann aus den Sprechmerkmalen auf die Persönlichkeit. Eine ausschließliche Validierung über die Berechnung solcher Korrelationen hält auch Rainer Bäcker, Leiter Managementdiagnostik bei der ifp Personalberatung in Köln für nicht ausreichend. „Dies ist letztlich eine Validierung ‚zweiter Ordnung‘, bei der man nie über die Validität des Persönlichkeitstests kommen kann“, erklärt der Diplom-Psychologe. Bäcker, der auch im wissenschaftlichen Beirat von Precire ist, will den Test daher ausschließlich bei Verhaltensanalysen einsetzen. „Das Besondere an Precire sei, dass man eine Vielzahl von Sprachmerkmalen gleichzeitig erfassen könne und zwar unvergleichlich mehr als dies durch ei-

ne reine Verhaltensbeobachtung bisher möglich war.“ Das ifp will den Test bei der Vertriebs- und Führungskommunikation nutzen und dabei zum Beispiel die Sprachmerkmale der erfolgreichsten Vertriebsmitarbeiter als Referenzgruppe nehmen. „Die Mitarbeiter bekommen dann Hinweise, wie sie überzeugender wirken können“, so Bäcker. Dass der Verkaufserfolg von Sprachmerkmalen abhängt, sei zwar eine Hypothese, aber „eine, für die es viele gute Gründe gibt“. Bei Precire handele es sich um einen vielversprechenden Ansatz, der „aber noch nicht durchgängig auf wissenschaftlich abgesichertem Boden steht“.

Das könne man zwar machen, erklärt Uwe P. Kanning, Psychologieprofessor an der Hochschule Osnabrück. „Das Problem bei dieser Art der externen Validierung sei jedoch, dass man nicht wisse, ob die erfassten Merkmale der erfolgreichen Mitarbeiter, auch tatsächlich mit deren Leistungsfähigkeit zusammenhängen. Wenn alle guten Mitarbeiter Schuhgröße 42 haben, ist es dennoch nicht sinnvoll, neue Mitarbeiter nach der Schuhgröße auszuwählen“, erklärt Kanning. „Das ist gewissermaßen eine illusorische Korrelation.“ Man komme daher nicht darum herum, auch nachzuweisen, dass die Schuhgröße (oder bestimmte Sprachmerkmale) direkt mit dem Erfolg zusammenhängen.

### Referenzkunden zurückhaltend

Bei der Zeitarbeitsfirma Randstad ist Precire bereits im Einsatz. „Bei Randstad nutzen wir Precire seit September 2014 in der Personalauswahl mit einem überzeugenden Ergebnis“, wird HR-Direktor Andreas Bolder auf der Psyware-Website zitiert. Auf Nachfrage erklärt Randstad, dass der Sprachtest nur bei der Suche von internen Mitarbeitern eingesetzt werde, die man „weniger nach Fachkompetenz, sondern eher nach Persönlichkeitsmerkmalen auswähle“. Als Kunden werden in einem FAZ.NET-Artikel vom 20. Mai 2015 auch die Accor Hotels genannt. Das System sei in eingegrenztem

Rahmen mit Führungskräften als Einzelprobelauf getestet worden, schreibt Eike Kraft, Pressesprecher bei den Accor Hotels. Es befinde sich heute in keiner Weise bei Accor Hotels im Einsatz und sei auch nicht für einen Einsatz geplant.

### Messverfahren nicht offengelegt

Angelika Braun, Professorin für Phonetik an der Universität Trier, sieht die Auswertungen von Precire kritisch. „Man kennt die Algorithmen nicht und das Messverfahren wird nicht offengelegt“, kritisiert Braun, die auch Leiterin des Fachbereichs Sprechererkennung, Tonträgerauswertung und Linguistische Textanalyse beim Bundeskriminalamt Wiesbaden war. „Zwar verweise Precire auf einige durchaus seriöse Studien, doch man wisse nicht, wie diese in der Software berücksichtigt werden. Zudem hätten sich die Forscher in den Studien stets mit bestimmten Fragestellungen beschäftigt und keine Gesamtuntersuchung gemacht, was das für die Persönlichkeit bedeutet. Bei Negationen heißt es etwa, dass sie speziell bei depressiven Fragestellungen „entscheidende Kategorien“ seien. Bei der erwähnten Studie geht es um „Linguistische Marker von depressiven Dynamiken in Selbsterzählungen“. Das hat jedoch nichts mit Personalauswahl zu tun. „Das Ganze basiert auf den beiden Grundannahmen, dass das stimmlich-sprachliche Verhalten konstant ist und es eine 1:1-Beziehung zwischen diesem Verhalten und Charaktereigenschaften gibt“, erklärt die Vizepräsidentin der International Society for the Phonetic Sciences. Doch beides treffe nicht zu. „Ob ich am Abend davor ein paar Gläser getrunken habe, erkälte bin oder rauche, hat alles Auswirkungen auf bestimmte Audio-Parameter“, erklärt Braun. „Die stimmlichen Merkmale sind daher ganz und gar nicht konstant innerhalb einer Person.“

Beim Blick auf meine Auswertung fallen ihr sachliche Fehler auf. So zähle man bei der „Basislinguistik“ im Deutschen aufgrund der oft langen Wortzusammen-

setzungen keine Wörter, sondern Silben. Auch werde die Kategorie „Sätze“ in der Linguistik bei gesprochener Sprache nicht akzeptiert, weil wir nicht mit Punkt und Komma sprechen. Stattdessen werte man Äußerungen aus, so die Professorin. Noch auffälliger wird es bei den „Audio-prosodischen Parametern“, also den Stimmeeigenschaften. Mein Maximalwert bei der Tonlage erscheine ihr ungewöhnlich hoch, das spreche für einen Messfehler. Auch die angegebene Definition von „Harmonizität“ als „das Verhältnis vom untersuchten Signal zu unerwünschtem Hintergrundrauschen“ kennt die Professorin nicht. In der Phonetik beschreibe „Harmonizität“ das Verhältnis von harmonischen und unharmonischen Anteilen in der Stimme. So haben zum Beispiel Raucher mehr unharmonische Anteile.

### Begrifflichkeiten unsauber verwendet

„Hier werden elementare Begrifflichkeiten nicht sauber verwendet“, sagt die Phonetik-Expertin. Verwundert ist sie auch bei der Erklärung zu „Perturbation“. Dort heißt es: „Dieses Zittern entsteht durch körperliche Phänomene, die auf Anspannung und/oder Unsicherheit hinweisen können.“ „Diese Interpretationen sind mir aus der Phonetik nicht bekannt“, sagt Braun. Zudem werde Perturbation nie bei fließender gesprochener Sprache erhoben, sondern man lasse die Person mehrere Sekunden einen Vokal halten mit exakt festgelegtem Abstand zum Mikro. „Um bei der Stimmanalyse zu aussagekräftigen Ergebnissen zu kommen, muss man sehr viel beachten“, erklärt Braun. Ein Interview per Telefon sei eine „gefährliche Signalbeschneidung“. Man wisse nicht mehr genau, was man misst. Das Precire-Modul zu den Stimmeeigenschaften sei daher für sie ohne erkennbare Aussagekraft. Die Wissenschaftlerin: „Ich würde mich als Bewerber nicht mit dieser Software analysieren lassen.“ ■

---

**BÄRBEL SCHWERTFEGER** arbeitet als freie Journalistin in München.

---

