



Universidad de Valladolid



PROGRAMA DE DOCTORADO EN ECONOMÍA

TESIS DOCTORAL:

**ORDINAL TREATMENT OF ORDERED QUALITATIVE
SCALES: ANALYSIS, METHODS AND APPLICATIONS**

**TRATAMIENTO ORDINAL DE LAS ESCALAS
CUALITATIVAS ORDENADAS: ANÁLISIS, MÉTODOS
Y APLICACIONES**

Presentada por Raquel González del Pozo para optar al
grado de Doctora por la Universidad de Valladolid

Dirigida por:

José Luis García Lapresta

ORDINAL TREATMENT OF ORDERED QUALITATIVE
SCALES: ANALYSIS, METHODS AND APPLICATIONS

TRATAMIENTO ORDINAL DE LAS ESCALAS
CUALITATIVAS ORDENADAS: ANÁLISIS, MÉTODOS
Y APLICACIONES

Raquel González del Pozo

Memoria presentada para optar al grado
de Doctora por la Universidad de Valladolid

Dirigida por:

José Luis García Lapresta

Valladolid, mayo de 2020

A Héctor, Javier
y a mis padres

Agradecimientos

Al intentar resumir lo que ha significado para mí la formación doctoral recibida durante estos años, me vienen a la memoria multitud de personas que han participado directa o indirectamente en esta tesis. Sin embargo, he de reconocer que la elaboración de este trabajo nunca hubiera sido posible sin el apoyo y la confianza de mi familia, la verdadera piedra angular de esta tesis.

Quiero empezar expresando mi más profundo agradecimiento a toda mi familia por su apoyo y su ánimo en todo momento, en especial a mi marido Javier, y a mi hijo Héctor, quienes me han acompañado y alentado en este camino y que no dudaron en estar a mi lado en la estancia de investigación que realicé en la Universidad de Coimbra. Gracias por su paciencia y su comprensión. Agradezco también el apoyo incansable de mis padres Benigno y María del Carmen, quienes me han inculcado la educación y valores necesarios para afrontar un desafío como este. Me gustaría también agradecer la ayuda que me han prestado mis suegros para conciliar mi vida laboral y familiar.

Desde el punto de vista académico, quiero comenzar expresando mi más sincera gratitud a mi director de tesis, José Luis García Lapresta, quien confió en mí desde un primer momento para la realización de esta tesis, y me alentó para adentrarme en este duro reto. Gracias por su tiempo, dedicación y consejos, sin los cuales no habría sido posible llegar hasta aquí.

Deseo dar las gracias a mis compañeros del grupo de investigación PRESAD por su acogida, ayuda, gentileza y disponibilidad para asesorarme en todas las ocasiones en las que he acudido a ellos. En particular, quiero resaltar la inestimable ayuda de David Pérez Román en el diseño de programas informáticos, sin los cuales no hubiera sido posible culminar este trabajo, así como los valiosos y sabios consejos de Miguel Martínez Panero.

Asimismo, quiero agradecer el apoyo recibido por parte de mi tutor Carlos

Rodríguez Palmero, y por los demás compañeros del Departamento de Economía Aplicada de la Universidad de Valladolid. Muchas gracias por vuestra ayuda y disponibilidad durante todos estos años que hemos compartido juntos.

Me gustaría agradecer también a Luís C. Dias y a los miembros del grupo de investigación CeBER de la Universidad de Coimbra, la acogida, sugerencias y consejos que me brindaron durante mi estancia.

No puedo olvidarme de mis compañeros de despacho, Verónica, Ziba y Sergio, quienes han estado siempre dispuestos a escucharme y me han ofrecido su ayuda cuando la he necesitado. Del mismo modo, también me gustaría tener palabras de agradecimiento para todos mis compañeros de travesía: Mafalda, Pablo, Jaime y demás doctorandos con los que he compartido muchas alegrías y experiencias enriquecedoras durante estos años.

Por último, quiero transmitir mi más sincera gratitud a mi amigo Íñigo, la persona que despertó en mí sin saberlo el deseo de iniciar mi carrera investigadora.

Raquel González del Pozo
Universidad de Valladolid
Mayo de 2020

Índice general

Agradecimientos	v
Estructura de la tesis doctoral	1
Introducción	3
Introduction	17
1 Metrizable ordinal proximity measures and their aggregation	31
2 Tratamiento ordinal de las escalas cualitativas utilizadas por el Centro de Investigaciones Sociológicas	33
3 An ordinal multi-criteria decision-making procedure under imprecise linguistic assessments.	35
4 Using different qualitative scales in a multi-criteria decision-making procedure	37
Conclusiones y trabajo futuro	39
Conclusions and further research	43
Bibliografía/Bibliography	47

Estructura de la tesis doctoral

Esta tesis doctoral se presenta bajo la modalidad de tesis por compendio de publicaciones, de acuerdo con la normativa vigente para la presentación y defensa de la tesis doctoral de la Universidad de Valladolid, aprobada por el Consejo de Gobierno el 3 de junio de 2016 (BOCyL nº 114 del 15 de junio de 2016). En ella se incluyen un total de cuatro artículos publicados en un medio de impacto, según los criterios de la ANECA para el área de conocimiento en la que se presenta la tesis.

Los artículos primero, tercero y cuarto están publicados en revistas indexadas en WoS-JCR y el segundo en SJR, por lo que cumplen con los requisitos establecidos por la Comisión Académica del Programa de Doctorado en Economía de la Universidad de Valladolid. A continuación se indican los artículos que conforman la tesis doctoral, las revistas donde están publicados y sus bases de indexación. Asimismo también se recoge la filiación de los coautores.

1. García-Lapresta, J.L., González del Pozo, R., Pérez-Román, D.: Metrizable ordinal proximity measures and their aggregation. *Information Sciences* 448-449, pp. 149-163, 2018.

Revista recogida en WoS-JCR 2018: 5,524 (JIF), Q1 (JIF Percentil: 94,516; 9/155). Computer Sciences, Information Systems.

2. González del Pozo, R., García-Lapresta, J.L.: Tratamiento ordinal de las escalas cualitativas utilizadas por el Centro de Investigaciones Sociológicas. *Revista de Métodos Cuantitativos para la Economía y la Empresa* 28, pp. 124-142, 2019.

Revista recogida en SJR 2019: 0,17 (JIF), Q3. Economics, Econometrics and Finance.

3. García-Lapresta, J.L., González del Pozo, R.: An ordinal multi-criteria decision-making procedure under imprecise linguistic assessments. *European Journal of Operational Research* 279, pp. 159-167, 2019.
Revista recogida en WoS-JCR 2018: 3,806 (JIF), Q1 (JIF Percentil: 85,119; 13/84). *Operations Research and Management Science*.
4. González del Pozo, R., Dias L.C, García-Lapresta, J.L.: Using different qualitative scales in a multi-criteria decision-making procedure. *Mathematics* 2020, 8(3), 458; doi:10.3390/math8030458.
Revista recogida en WoS-JCR 2018: 1,105 (JIF), Q1 (JIF Percentil: 76.274; 75/314). *Mathematics*.

Filiación de los coautores

Luís C. Dias

Catedrático de Métodos Científicos de Gestão. Faculdade de Economia. Universidade de Coimbra. Portugal.

José Luis García Lapresta

Catedrático de Universidad de Métodos Cuantitativos para la Economía y la Empresa. Departamento de Economía Aplicada. Universidad de Valladolid.

David Pérez Román

Profesor Titular de Universidad de Organización de Empresas. Departamento de Organización de Empresas y Comercialización e Investigación de Mercados. Universidad de Valladolid.

Introducción

Los procedimientos de toma de decisiones juegan un papel fundamental en la vida de los seres humanos y organizaciones. En multitud de ocasiones los individuos manifiestan sus opiniones y preferencias sobre un conjunto de alternativas respecto a uno o varios criterios con el fin de determinar una decisión colectiva: procesos de control de calidad, encuestas de satisfacción y calidad de vida, análisis de sistemas de evaluación de desempeño, elección de representantes políticos, evaluación de la actividad docente, etc.

Dado un conjunto de alternativas, identificar el tipo de problema es el primer paso al que se enfrentan los individuos a la hora de tomar una decisión. Autores como Roy [92, 93], Doumpos y Zopounidis [31], Kadzinski et al. [63] y Figueira et al. [35], entre otros, distinguen cuatro tipos de problemas de decisión multi-criterio, considerándose los tres primeros como los más importantes: elección (elegir la mejor alternativa), clasificación (agrupar las alternativas en categorías predefinidas), ranking (ordenar las alternativas de acuerdo con un orden de preferencia decreciente), y descripción (describir el problema y sus consecuencias).

En los numerosos procedimientos de toma de decisiones en los que las opiniones y valoraciones de los individuos no pueden ser evaluadas mediante valores numéricos precisos, es más apropiado adoptar un enfoque lingüístico. Por ejemplo, cuando se evalúan determinados aspectos sensoriales y subjetivos como la salud autopercebida, sentimientos personales o gustos y preferencias. En casos como los anteriores es frecuente manejar información de tipo cualitativo la cual puede venir expresada mediante palabras, términos lingüísticos u oraciones pertenecientes a un lenguaje natural o artificial (véanse Herrera et al. [55], Delgado et al. [26] y Herrera et al. [53], entre otros).

Las palabras son el medio más utilizado por los individuos para dar a conocer sus opiniones y preferencias y además son el principal competidor de la probabilidad a

la hora de representar la incertidumbre (véanse Von Winterfeldt y Edwards [104] y Gambará [42]). En este sentido autores como Zimmer [115, 116] y Windschitl y Wells [109] apuntan a que la comunicación verbal es más natural que la numérica, debido a que la capacidad de entender y utilizar adecuadamente el lenguaje se desarrolla antes que el dominio y la comprensión de los números.

La mejor comprensión de la información verbal y la dificultad para representar mediante valores numéricos exactos las opiniones de los individuos en situaciones de vaguedad e imprecisión, son algunas de las razones por las que las expresiones verbales son preferidas a las numéricas (véanse Beyth-Marom [11], Zimmer [115], Wallsten et al. [105], Wallsten et al. [107], Budescu y Wallsten [17], Budescu et al. [18], Teigen [99], Jaffe-Katz et al. [60], Wallsten et al. [106] y Dhami y Mandel [27], entre otros). Esta incertidumbre propia del comportamiento humano ha sido abordada en diversos procedimientos de toma de decisiones (véanse Novák [81], Chen y Hwang [21], Pawlak y Slowinski [85] y Pawlak [84], entre otros) mediante conjuntos difusos (*fuzzy sets*) (véase Zadeh [111, 112]), conjuntos aproximados (*rough sets*) (véase Pawlak [83]) o a través de la computación con palabras (*computing with words*) (véanse Zadeh [113, 114] y Herrera et al. [52]).

Debido a la preferencia de los seres humanos por las palabras y a que en ocasiones es necesario manejar información que no puede ser valorada cuantitativamente, es frecuente el uso de escalas cualitativas formadas por un determinado número de etiquetas o cuantificadores lingüísticos. A a estas etiquetas se les conoce comúnmente con el nombre de categorías de respuesta (lingüística) dentro del ámbito de las Ciencias Sociales: Economía, Marketing, Política, Psicología o Sociología. No obstante, en el contexto de la lógica difusa suele utilizarse la expresión *término lingüístico* para hacer referencia a las etiquetas lingüísticas de las escalas cualitativas (véase Zadeh [112]).

Las escalas cualitativas o categóricas permiten clasificar elementos de acuerdo con ciertas cualidades o características: sexo, raza, nivel de estudios, clases sociales, etc. Estas escalas son más apropiadas que las numéricas para conocer las opiniones, preferencias o necesidades de los individuos, ya que las categorías de respuesta lingüística precisan un menor tiempo de procesamiento de la información y proporcionan a los cuestionarios mejores propiedades psicométricas (véanse Prieto [87] y Durán et al. [32], entre otros).

Estas escalas cualitativas pueden dividirse en dos tipos: nominales y ordinales

(véase Stevens [97]). Las nominales son aquellas que clasifican los elementos en categorías mutuamente excluyentes y exhaustivas. Así, la actividad profesional podría modelizarse a través de una escala nominal como la siguiente: {“profesor”, “abogado”, “médico”, “cocinero”}. Por el contrario, las escalas cualitativas ordenadas u ordinales organizan los elementos siguiendo determinados criterios de orden o posición. Por ejemplo: las medallas de una prueba deportiva: {“bronce”, “plata”, “oro”} o el nivel de estudios de una población: {“sin estudios”, “estudios primarios”, “estudios secundarios”, “estudios superiores”}.

Las escalas cualitativas ordenadas son utilizadas frecuentemente en los cuestionarios empleados por multitud de organismos e instituciones para la realización de encuestas. La Oficina Estadística de la Unión Europea (EUROSTAT) utiliza diversas escalas cualitativas como las siguientes: {“con gran dificultad”, “con dificultad”, “con alguna dificultad”, “con bastante facilidad”, “con facilidad”, “muy fácilmente”} (EU-SILC HS120) y {“carga pesada”, “carga razonable”, “ninguna carga”} (EU-SILC HS140) en los cuestionarios de la Encuesta Europea de Ingresos y Condiciones de Vida. Asimismo, el Centro de Investigaciones Pew Research Center analizó en 2016 las opiniones de los estadounidenses acerca de los nueve candidatos a la presidencia de los Estados Unidos a través de una escala cualitativa de 5 términos: {“terrible”, “poor”, “average”, “good”, “great”}.

Asimismo, Balinski y Laraki [6, 7, 8, 9] han propuesto un nuevo sistema de votación denominado *Majority Judgement* que también utiliza escalas cualitativas como la siguiente: {“to reject”, “poor”, “acceptable”, “good”, “very good”, “excellent”}. En este procedimiento se pide a los participantes que evalúen un conjunto de alternativas a través de una escala formada por términos lingüísticos. Durante las elecciones presidenciales francesas de abril de 2007 se realizó un análisis experimental de este sistema de votación, pidiendo a los participantes que otorgaran un término lingüístico de la escala cualitativa anterior a cada uno de los candidatos presidenciales (véase Balinski y Laraki [7]).

A pesar de que existe una amplia literatura en la que se discute el número de términos o categorías de respuesta que deben incluirse en escalas cualitativas, aún no se ha alcanzado una conclusión definitiva (véanse Lissitz y Green [69], Cox [25], Oaster [82], Preston y Colman [86] y Matas [78], entre otros). En general, las evidencias empíricas han demostrado que se debe evitar la incorporación de pocas categorías de respuesta, ya que esto puede afectar a la precisión y a la validez de

las escalas haciendo que los individuos no sean capaces de expresar sus opiniones correctamente cuando responden a los cuestionarios. Por otra parte, tampoco es conveniente introducir un número excesivo de categorías de respuesta debido a la limitada capacidad de los individuos para procesar y retener información. Por estas razones, la mayor parte de las escalas cualitativas suelen emplear de cuatro a siete categorías de respuesta (véanse Miller [80] y Lozano et al. [73]).

En lo referente al diseño de encuestas y cuestionarios, la decisión de utilizar un número par o impar de categorías de respuesta también ha sido objeto de debate (véanse Garland [48], Johns [62], Baka et al. [5] y Díaz de Rada [29], entre otros). Generalmente, se utilizan escalas cualitativas formadas por un número impar de categorías de respuesta. Estas escalas suelen ser equilibradas, es decir, presentan el mismo número de términos favorables y desfavorables, separados por un término central que permite a los individuos posicionarse de una forma neutral, e.g. {“muy en desacuerdo”, “desacuerdo”, “indiferente”, “de acuerdo”, “muy de acuerdo”}. No obstante, cuando se desea forzar a los individuos a posicionarse en algún punto concreto, pueden emplearse escalas cualitativas formadas por un número par de categorías de respuesta: {“nada”, “poco”, “bastante”, “mucho”}. Esta escala cualitativa de cuatro términos es utilizada frecuentemente en los barómetros del Centro de Investigaciones Sociológicas (CIS).

Las escalas de respuesta tipo Likert [68] sirven para medir actitudes, conocimientos, opiniones y otros aspectos cualitativos. Se trata de un tipo de escala aditiva cuyo fin es conocer el grado de acuerdo o desacuerdo de los individuos en relación con una serie de elementos, afirmaciones o ítems. Normalmente, este tipo de escalas son equilibradas e incluyen un término de respuesta neutral (véanse Lalla et al. [66] y Matas [78], entre otros). Por ejemplo, en la investigación desarrollada por Caballero Rodríguez [19] se emplea la siguiente escala Likert de 5 puntos: {1=“muy bajo”, 2=“bajo”, 3=“medio”, 4=“alto”, 5=“muy alto”} para conocer el nivel de satisfacción del profesorado respecto a ocho afirmaciones relacionadas con el acceso y la promoción en la universidad.

Como se puede ver en el ejemplo anterior, una de las características de las escalas tipo Likert es que utilizan la misma escala cualitativa para valorar todos los elementos o afirmaciones, los cuales suelen tener asignado un mismo peso o importancia dentro de la escala. Finalmente, una vez contestados todos los elementos se calcula la suma o el promedio de las respuestas a fin de determinar la posición de los

individuos con relación a la variable objeto de estudio.

Las escalas cualitativas formadas por términos lingüísticos son una de las herramientas más utilizadas en la investigación cualitativa. No obstante, algunas de las ventajas de estas escalas se ven contrarrestadas por el delicado análisis estadístico y matemático que precisan, puesto que no hay criterios definidos acerca de qué técnicas o procedimientos deben emplearse. Algunos de los métodos propuestos para resolver este problema son: la asignación de valores numéricos o puntuaciones a cada uno de los términos lingüísticos de las escalas (véanse Yates [110], Calatrava Requena [20], Roberts [88] y Fielding [34], entre otros), el uso de herramientas enmarcadas dentro de la lógica difusa como los modelos basados en números difusos (*fuzzy numbers*) y funciones de pertenencia (véanse Lalla et al. [66] y Gil y González-Rodríguez [49], entre otros), y modelos de representación lingüística, como el *modelo 2-tuplas* introducido por Herrera y Martínez (véanse Herrera y Martínez [56], y Martínez et al. [75], entre otros). Este modelo ha sido aplicado a contextos en los que se maneja información lingüística procedente de múltiples escalas cualitativas (véase Herrera et al. [53]), así como en contextos en los que se utilizan escalas cualitativas no equilibradas, es decir, escalas en las que los términos lingüísticos no se distribuyen de una manera simétrica alrededor de un término central (véase Herrera et al. [54]).

El *modelo 2-tuplas* toma como base el concepto de translación simbólica para representar la información lingüística. En este modelo se representa mediante un par de valores denominado 2-tupla lingüística el resultado obtenido tras realizar operaciones de agregación con los índices de los términos lingüísticos. El primer componente de la tupla representa los índices de los términos lingüísticos, mientras que el segundo indica la distancia existente desde el resultado obtenido al índice del término lingüístico más cercano. Por tanto, este procedimiento equivale matemáticamente a trabajar con valores numéricos, una vez que a cada término lingüístico se le ha asociado con un número entero (véase García-Lapresta [43]).

Otras propuestas introducidas para manejar datos cualitativos son: el método de análisis Redit en el que las puntuaciones o “ridits” se obtienen teniendo en cuenta la frecuencia relativa de las categorías de respuesta (véase Bross [16]), y las funciones de puntuación media condicional que sugieren una función de puntuación basada en el supuesto de que los datos son generados por distribuciones subyacentes (véanse Fielding [34] y Chen y Wang [22]).

La asignación de valores numéricos o puntuaciones a los términos lingüísticos

de las escalas es el método más empleado para analizar información procedente de escalas cualitativas. Sin embargo, la principal crítica que puede hacerse a este método es la subjetividad que aporta al utilizar valores numéricos para representar información cualitativa (véase Calatrava Requena [20]).

Por otra parte, el uso de valores numéricos puede dar a entender que existe algún tipo de relación de proporcionalidad entre las categorías de respuesta. No obstante, esta relación no es cierta, puesto que en las escalas ordinales las diferencias entre los valores numéricos no miden ninguna magnitud que tenga sentido (véanse Grande Esteban [50] y Agresti [3], entre otros). Así, por ejemplo, al asignar puntuaciones equidistantes a los términos lingüísticos de la escala empleada por el CIS: {“nada”, “poco”, “bastante”, “mucho”}, podría considerarse que los términos lingüísticos están equiespaciados, a pesar de que no tenemos ninguna evidencia de que la diferencia cualitativa entre “mucho” y “bastante” pueda ser representada por el mismo valor numérico que la diferencia entre “poco” y “bastante” (véanse Calatrava Requena [20], Silva Aycaguer [96], Fleiss et al. [36], López-González [72], Chen y Wang [22] y Martínez Martín [77], entre otros).

Es fundamental tener en cuenta que, dado que en las escalas cualitativas ordenadas los valores numéricos no representan una relación cuantitativa entre las distintas categorías de respuesta, no deben realizarse operaciones aritméticas con estos números. En esta línea Stevens [97, p. 679] apunta:

Los estadísticos comunes como la media y desviación estándar no deben usarse en escalas ordinales, ya que estos estadísticos implican el conocimiento de algo más que el mero orden jerárquico relativo de los datos. Por otro lado, este uso “ilegal” de la Estadística está avalado por una suerte de ley pragmática: en numerosas ocasiones nos lleva a resultados fructíferos. Pero, si bien la prohibición de este procedimiento probablemente no sirva para nada, es apropiado señalar que las medias y las desviaciones estándar calculadas en una escala ordinal son erróneas en la medida en que los intervalos sucesivos en la escala son de tamaño desigual.

De igual forma, otros autores como Gardner [47], Roberts [88], Marcus-Roberts y Roberts [74], Knapp [64] y Franceschini [37], entre otros, también han cuestionado ampliamente el uso generalizado de técnicas y procedimientos estadísticos para el

tratamiento y análisis de datos cualitativos.

A pesar de los problemas que presenta la transformación de la información cualitativa en cuantitativa, muchos organismos nacionales e internacionales continúan aplicando estas técnicas. Esta praxis poco recomendable en el manejo de las escalas cualitativas puede conducir en ocasiones a una interpretación distorsionada de la información, ya que pueden obtenerse resultados diferentes en función de cómo sean codificadas las categorías de respuesta de las escalas (véanse Stevens [97, 98], Merbitz et al. [79], Saaty [94], Franceschini [37], Franceschini et al. [38] y Gadrich et al. [41], entre otros). Asimismo, Siegel [95, p. 3] también señala algunos de los inconvenientes del uso de puntuaciones numéricas en las pruebas estadísticas paramétricas:

Cuando se realizan operaciones aritméticas con puntuaciones que no son verdaderamente numéricas, se ocasionan naturalmente deformaciones de los datos y se menoscaba el valor de las conclusiones de la prueba.

Con el fin de superar estos inconvenientes y analizar los datos ordinales de una forma adecuada, se han propuesto en la literatura diversos procedimientos basados en criterios de dominancia y en las funciones de distribución de las variables categóricas. Así, Blair y Lacy [12] proponen una medida de dispersión basada en las frecuencias relativas acumuladas. En el trabajo de Martínez Martín [77] se proponen dos generalizaciones de medidas de dispersión para variables ordinales basadas en las frecuencias de las categorías de respuesta, ponderadas por una medida de discrepancia. Franceschini et al. [39] introducen un método de control de calidad que utiliza información procedente de escalas ordinales. Este procedimiento está basado en criterios de dominancia de clasificación y dispersión (veáanse Franceschini y Romano [40] y Franceschini et al. [38]), y permite realizar análisis estadísticos sin utilizar operadores tradicionales como la media o la varianza muestral ni codificar de forma arbitraria las categorías de respuesta de las escalas. Bashkansky y Gadrich [10] también abordan los procesos de control de calidad a través de un método conocido como *Proportion Ratio and Dispersion* (PR&D) que genera escalas de calidad teniendo en cuenta las frecuencias relativas de las categorías y su dispersión. Por otra parte, Herrero y Villar [58] manejan los datos de tipo categórico mediante un procedimiento basado en la distribución de los datos y en las relaciones de dominancia por pares.

Otro factor que puede influir a la hora de analizar y manejar datos cualitativos

es la forma en la que los individuos perciben este tipo de escalas. Habitualmente, las escalas cualitativas ordenadas son consideradas como uniformes, es decir, existe la misma proximidad entre todos los términos lingüísticos consecutivos que las componen los cuales se distribuyen de forma simétrica alrededor de un término o categoría de respuesta central. Por ejemplo: {“muy malo”, “malo”, “regular”, “bien”, “muy bien”}. No obstante en ocasiones, el sentido común, el valor semántico de las palabras, las diferencias culturales o incluso la propia personalidad de los individuos pueden hacer que las escalas cualitativas sean percibidas como no uniformes (véanse Silva Aycaguer [96], Clarke [23], Javaras y Ripley [61], Hofmans et al. [59] y Martínez García y Martínez Caro [76], entre otros). Una escala cualitativa ordenada puede ser considerada no uniforme cuando los individuos aprecian proximidades psicológicas distintas entre los términos lingüísticos consecutivos de la misma. Así pues, la escala cualitativa del CIS mencionada anteriormente puede ser considerada como no uniforme si los individuos perciben que el término lingüístico “bastante” está más próximo a “mucho” que a “poco”, o si el término “poco” está más próximo a “nada” que a “bastante”.

En el contexto de las escalas cualitativas no uniformes, no es apropiado asignar valores numéricos o puntuaciones a los términos lingüísticos, ya que estos pueden no reflejar adecuadamente la percepción que tienen los individuos acerca de las proximidades entre los términos de las escalas.

En esta tesis, para manejar las escalas cualitativas no uniformes de una forma puramente ordinal, se utiliza el concepto de *medida de proximidad ordinal* introducido por García-Lapresta y Pérez-Román [45]. Las medidas de proximidad ordinal recogen a través de un conjunto de grados de proximidad la información sobre cómo perciben los individuos las proximidades entre los términos de las escalas. Es importante señalar que estos grados de proximidad no denotan números. Son objetos abstractos que representan diferentes proximidades.

El objetivo principal de esta tesis es el análisis, extensión y aplicación de las medidas de proximidad ordinal a diferentes contextos. Para ello, se han diseñado diversos procedimientos para generar y agregar medidas de proximidad ordinal, y se han propuesto nuevos modelos de toma de decisiones multi-criterio en los que se utilizan estas medidas.

En el **capítulo 1** se analizan los procedimientos de generación y agregación de medidas de proximidad ordinal. En él se introduce el concepto de medida de

proximidad ordinal metrizable: medidas de proximidad ordinal en las que las comparaciones ordinales entre los términos lingüísticos pueden representarse a través de una métrica lineal sobre la escala cualitativa.

Asimismo, en este capítulo se presenta un algoritmo que evita problemas e inconsistencias a la hora de generar medidas de proximidad ordinal metrizable en escalas cualitativas de cuatro términos. Este algoritmo permite obtener para cada experto su medida de proximidad ordinal metrizable a través de una secuencia apropiada de preguntas sobre las proximidades entre los términos de la escala. Dado que los expertos pueden percibir proximidades psicológicas distintas entre los términos de una escala cualitativa ordenada, es posible obtener diferentes medidas de proximidad ordinal metrizable al aplicar el algoritmo. En estos casos surge un problema de agregación de las opiniones de los expertos el cual puede enmarcarse dentro de la Teoría de la Elección Social (véase List [71]).

Desde el punto de vista de la Teoría de Agregación de Juicios (*Judgment Aggregation*) (véanse Dietrich y List [30], List [70] y Grossi y Pigozzi [51], entre otros) se han analizado y propuesto diversos procedimientos que buscan obtener un juicio colectivo que refleje lo más fielmente posible las opiniones de un grupo de expertos. No obstante, es importante destacar que la elección del procedimiento de agregación no es algo trivial, puesto que al aplicar diferentes sistemas de votación pueden obtenerse resultados distintos e incluso inconsistencias.

Con objeto de evitar problemas e inconsistencias al agregar las opiniones de los expertos, se propone en este capítulo un procedimiento de agregación basado en métricas ponderadas. Este procedimiento considera que la medida de proximidad ordinal metrizable colectiva es aquella que minimiza la suma de distancias (o suma de distancias al cuadrado) a las medidas de proximidad ordinal metrizable de los expertos.

El capítulo también incluye algunos ejemplos ilustrativos del funcionamiento del algoritmo y del procedimiento de agregación.

En el **capítulo 2** se presenta una aplicación del sistema de votación introducido por García-Lapresta y Pérez-Román [46] a un conjunto de datos tomados del Barómetro del CIS.

En este capítulo se discute la metodología de agrupación de varias categorías de respuesta que realiza el CIS y otros organismos nacionales e internacionales como el Pew Research Center. Dicha práctica consiste en agrupar los resultados de determi-

nadas categorías de respuesta (por ejemplo “mucho” y “bastante” o “very good” y “somewhat good”), descartándose el resto de categorías de respuesta. La aplicación de esta metodología está asignando implícitamente un mismo valor numérico a las categorías de respuesta agrupadas y una puntuación nula a las categorías omitidas, por lo que se está produciendo una pérdida importante de la información cualitativa de las escalas.

A fin de evitar esta pérdida de información y considerar todas las categorías de respuesta de las escalas cualitativas, se ha ordenado un conjunto de alternativas procedentes del Barómetro del CIS a través del sistema de votación propuesto por García-Lapresta y Pérez-Román [46]. Se trata de un procedimiento ordinal en el que las alternativas se ordenan teniendo en cuenta el par de medianas de grados de proximidad ordinal. Para cada una de las alternativas consideradas, estos grados de proximidad ordinal se obtienen de la comparación entre la máxima valoración cualitativa de la escala y las valoraciones otorgadas por los agentes.

El sistema de votación *Majority Judgment* (véase Balinski y Laraki [8]) utiliza la mediana como elemento representativo de las valoraciones de los individuos. En particular, este sistema toma la mediana como valoración global cuando el número de valoraciones es impar, y la mediana baja cuando el número es par. No obstante, existen algunas diferencias reseñables entre el sistema de votación *Majority Judgment* y el procedimiento aplicado en este capítulo. En el sistema de votación introducido por García-Lapresta y Pérez-Román [46] se tiene en cuenta cómo perciben los individuos las proximidades entre los términos de la escalas, aspecto no contemplado en el sistema de votación *Majority Judgment*. Por otra parte, en el caso de que el número de valoraciones sea par puede producirse una pérdida importante de información si se toma únicamente la valoración cualitativa que ocupa la posición de la mediana baja. Por esta razón y para evitar esta pérdida de información, el sistema aplicado en este capítulo considera las dos medianas de los grados de proximidad.

El sistema de votación propuesto por García-Lapresta y Pérez-Román [46] también introduce un proceso secuencial de desempate basado en Balinski y Laraki [6, 8]. Dicho proceso consiste en eliminar el par de medianas de grados de proximidad de las alternativas que están empatadas y seleccionar un nuevo par entre los grados de proximidad ordinal restantes. El proceso de desempate se repite las veces que sea necesario hasta romper todos los empates y ordenar todas las alternativas.

En el **capítulo 3** se introduce un procedimiento ordinal de toma de decisio-

nes multi-criterio. Este sistema es una extensión del procedimiento propuesto por García-Lapresta y González del Pozo [44] y del sistema de votación introducido por García-Lapresta y Pérez-Román [46].

En concreto, el procedimiento propuesto ordena un conjunto de alternativas valoradas por un grupo de agentes respecto a múltiples criterios. Para realizar las valoraciones, los agentes pueden utilizar una escala cualitativa no necesariamente uniforme.

Dado que el número de categorías de respuesta de las escalas es limitado, los agentes pueden dudar acerca de qué término lingüístico asignar a cada alternativa. Una manera de integrar la imprecisión en los procedimientos de toma de decisiones consiste en permitir a los agentes asignar varios términos lingüísticos consecutivos. A este respecto, existen en la literatura varios trabajos en los que se contempla dicha posibilidad (véanse Torra [102], Agell et al. [1], Rodríguez et al. [89], Agell et al. [2], Falcó et al. [33], Rodríguez et al. [90] y Wang et al. [108], entre otros).

El procedimiento introducido en este capítulo también incorpora la posibilidad de duda, permitiendo a los agentes valorar las alternativas a través de uno o dos términos lingüísticos consecutivos cuando no estén seguros de qué término otorgar. Para ello se introduce una extensión de las medidas de proximidad ordinal basada en los *espacios de órdenes de magnitud absoluta* introducidos por Travé-Massuyès y Piera [103], y en las extensiones de Falcó et al. [33] y Roselló et al. [91], entre otros.

Para representar en el procedimiento el peso de cada criterio, se propone replicar las valoraciones cualitativas de las alternativas las veces que sea necesario, hasta lograr que dichas replicaciones reflejen las ponderaciones de los criterios. Balinski y Laraki [8] también utilizan replicaciones en su propuesta de extensión del sistema de votación *Majority Judgment* a problemas multi-criterio. No obstante, es importante mencionar de nuevo que el sistema de votación *Majority Judgment* no considera la no uniformidad de las escalas cualitativas y además, no permite a los agentes utilizar más de un término lingüístico en sus valoraciones.

Finalmente, en este capítulo se presenta una aplicación del procedimiento propuesto para ordenar un conjunto de vinos teniendo en cuenta los aspectos sensoriales más importantes de una cata: apariencia, olor y sabor, así como la impresión global de cada uno de los vinos.

En el sistema de votación y en el procedimiento ordinal considerados en los capítulos anteriores se utiliza siempre la misma escala cualitativa ordenada para

valorar las alternativas. No obstante, en algunas situaciones se requiere a los agentes que manifiesten sus valoraciones a través de diversas escalas cualitativas formadas por un número distinto de términos lingüísticos (véanse Herrera et al. [53], Herrera et al. [57] y de Andrés et al. [4], entre otros).

En el **capítulo 4** se propone un procedimiento de toma de decisiones multi-criterio que maneja información procedente de diferentes escalas cualitativas. Dicho procedimiento pospone en todo lo posible el uso de valores numéricos con el fin de preservar y respetar la información ordinal de las escalas. El procedimiento propuesto permite a los agentes evaluar cada criterio a través de una escala cualitativa específica. Estas escalas cualitativas ordenadas pueden ser consideradas como no uniformes y estar formadas por un número distinto de términos lingüísticos.

Siguiendo los enfoques de la regla de Copeland y del método de decisión multi-criterio *PROMETHEE II* (véanse Copeland [24], Brans [14], Bouyssou [13] y Brans y De Smet [15]), el procedimiento propuesto genera un ranking de alternativas basado en un principio intuitivo de compensación entre las ventajas y desventajas de cada alternativa con relación a sus oponentes.

Para mostrar el funcionamiento del procedimiento multi-criterio se presenta un ejemplo ilustrativo en el que se introduce un enfoque estocástico para evaluar la robustez de los resultados obtenidos (véanse Lahdelma et al. [65], Tervonen y Lahdelma [101], Tervonen y Figueira [100] y Dias y Vetschera [28], entre otros). Asimismo, en este capítulo también se incluye una comparación del procedimiento propuesto con los métodos *SMAA-O* (véase Lahdelma [65]) y *ZAPROS III* (véase [67]).

Por último, la tesis finaliza con un capítulo de **conclusiones y trabajo futuro** en el que se resumen las principales contribuciones de esta tesis y se presentan algunas posibles líneas de investigación.

Asimismo, a medida que esta tesis se iba desarrollando, las propuestas y resultados obtenidos se han presentado en varios congresos y reuniones científicas, así como en diversos seminarios de investigación detallados a continuación:

Congresos y reuniones científicas nacionales:

- ESTYLF 2018. XIX Congreso Español sobre Tecnologías y Lógica Fuzzy. Granada, 2018.

- SEIO 2018. XXXVII Congreso Nacional de Estadística e Investigación Operativa y XI Jornadas de Estadística Pública. Oviedo, 2018.
- X Encuentro Interuniversitario de Investigadores en Decisiones Públicas. Pamplona, 2017.
- V Jornadas de Trabajo sobre Sistemas de Votación. Valladolid, 2017.
- ESTYLF 2016. XVIII Congreso Español sobre Tecnologías y Lógica Fuzzy. San Sebastián, 2016.

Congresos y reuniones científicas internacionales:

- DSRS-Turing 2019. 1st International Alan Turing Conference on Decision Support and Recommender Systems. Londres (Reino Unido), 2019.
- EURO 2019. 30th European Conference on Operational Research. Dublín (Irlanda), 2019.
- ASMDA 2019. 18th Applied Stochastic Models and Data Analysis International Conference. Florencia (Italia), 2019.
- EMPG 2018. Meeting of the European Mathematical Psychology Group. Génova (Italia), 2018.
- EWG/MCDA 2018. 87th Meeting of the European Working Group on Multi-criteria Decision Aiding. Delft (Países Bajos), 2018.
- MOPGP 2017. 12th International Conference on Multiple Objective Programming and Goal Programming. Metz (Francia), 2017.
- III International Workshop on Proximity Data, Multivariate Analysis and Classification. Valladolid (España), 2017.
- EUSFLAT 2017. The 10th Conference of the European Society for Fuzzy Logic and Technology. Varsovia (Polonia), 2017.
- IFORS 2017. 21st Conference of the International Federation of Operational Research Societies. Quebec City (Canadá), 2017.

- MCDM 2017. 24th International Conference on Multiple Criteria Decision Making. Ottawa (Canadá), 2017.
- Workshop on Future Directions in Computational Social Choice. Budapest (Hungría), 2016.

Seminarios de investigación

- Seminario de investigación organizado por el Centro de Estudos de Gestão do Instituto Superior Técnico (CEG-IST). Instituto Superior Técnico (IST). Universidade de Lisboa. Lisboa (Portugal), 2019.
- Seminario de investigación organizado por el Centre for Business and Economics Research (CeBER). Faculdade de Economia. Universidade de Coimbra. Coimbra (Portugal), 2018.
- Seminario de investigación organizado por el Grupo de Investigación de Ciencias Sociales Aplicadas. Facultad de Ciencias Económicas y Empresariales. Universidad de Valladolid. Valladolid (España), 2018.
- Seminario de investigación organizado por el Programa de Doctorado en Economía de la Empresa en colaboración con el Programa de Doctorado en Economía. Facultad de Ciencias Económicas y Empresariales. Universidad de Valladolid. Valladolid (España), 2018.

Introduction

Decision-making procedures play an important role in the lives of people and organizations. In many cases, individuals express their opinions and preferences over a set of alternatives in relation to one or more criteria in order to reach a collective decision: quality control procedures, satisfaction and quality of life surveys, analysis of performance appraisal processes, political elections, evaluation of teaching activity, etc.

Given a set of alternatives, the first step faced by individuals when making a decision is identifying the type of problem. Authors such as Roy [92, 93], Doumpos and Zopounidis [31], Kadzinski et al. [63] and Figueira et al. [35], among others, distinguish four different types of multiple criteria decision-making problems, considering the first three to be the most important: choice (choose the best alternative), sorting (sort the alternatives into predefined categories), ranking (rank the alternatives in decreasing order of preference), and description (describe the problem and its consequences).

In many decision-making procedures where individuals' opinions and assessments cannot be evaluated by precise numerical values, it is more appropriate to adopt a linguistic approach. For example, when some sensory and subjective aspects are evaluated: self-rated health, personal feelings or tastes and preferences. In these cases, it is common to use qualitative information that can be expressed in words, linguistic terms or sentences belonging to a natural or artificial language (see Herrera et al. [55], Delgado et al. [26] and Herrera et al. [53], among others).

Words are the most widely used way in which individuals express their opinions and preferences, and also the main alternative to probability when it comes to representing uncertainty (see Von Winterfeldt and Edwards [104] and Gambara [42]). Authors such as Zimmer [115, 116] and Windschitl and Wells [109] point out that verbal communication is more natural than its numerical counterpart, because the

ability to comprehend and employ language properly emerged long before people could understand and use numbers.

The greater comprehension of verbal information and the difficulty of representing individuals' opinions through exact numerical values in situations of vagueness and imprecision are just two of the reasons why verbal expressions are preferred to numerical ones (see Beyth-Marom [11], Zimmer [115], Wallsten et al. [105], Wallsten et al. [107], Budescu and Wallsten [17], Budescu et al. [18], Teigen [99], Jaffe-Katz et al. [60], Wallsten et al. [106] and Dhimi and Mandel [27], among others). This uncertainty, typical of human behavior, has been taken into account in various decision-making procedures (see Novák [81], Chen and Hwang [21], Pawlak and Slowinski [85] and Pawlak [84], among others) through *fuzzy sets* (see Zadeh [111, 112]), *rough sets* (see Pawlak [83]) or by means of *computing with words* (see Zadeh [113, 114] and Herrera et al. [52]).

Due to the human preference for words, and the fact that there is sometimes a need to manage information that cannot be assessed quantitatively, it is common to use qualitative scales formed by a certain number of linguistic labels or quantifiers. In the Social Sciences, these labels are commonly referred to as (linguistic) response categories. This is true for Economics, Marketing, Politics, Psychology and Sociology. Nevertheless, in the context of fuzzy logic the expression *linguistic term* is commonly used to refer to the linguistic labels used in qualitative scales (see Zadeh [112]).

Qualitative or categorical scales allow elements to be classified according to given qualities or characteristics: sex, race, level of studies, social classes, etc. These scales are more appropriate than numerical ones when identifying the opinions, preferences or needs of individuals, since linguistic response categories require less time to process the information and give questionnaires better psychometric properties (see Prieto [87] and Durán et al. [32], among others).

These qualitative scales can be divided into two different types: nominal and ordinal (see Stevens [97]). Nominal scales classify the elements into mutually exclusive and exhaustive categories. For instance, a person's profession could be modeled using a nominal scale such as: {"teacher", "lawyer", "doctor", "chef"}. In contrast, in ordinal or ordered qualitative scales the elements are classified following certain criteria relating to order or position. For example: the medals in a sporting event: {"bronze", "silver", "gold"} or people's educational levels: {"no schooling",

“primary studies”, “secondary studies”, “university studies”}.

Ordered qualitative scales are frequently used in questionnaires by many organizations and institutions that conduct surveys. The Statistical Office of the European Union (EUROSTAT) uses several ordered qualitative scales such as: {“with great difficulty”, “with difficulty”, “with some difficulty”, “fairly easily”, “easily”, “very easily”} (EU-SILC HS120) and {“a heavy burden”, “somewhat a burden”, “not burden at all”} (EU-SILC HS140) in questionnaires for European Union Statistics on Income and Living Conditions. Likewise, in 2016, the Pew Research Center analyzed the opinions of Americans about the nine United States presidential candidates through a 5-term ordered qualitative scale: {“terrible”, “poor”, “average”, “good”, “great”}.

In addition, Balinski and Laraki [6, 7, 8, 9] have proposed a new voting system called *Majority Judgment* which also employs ordered qualitative scales such as: {“to reject”, “poor”, “acceptable”, “good”, “very good”, “excellent”}. In this procedure, participants are asked to evaluate a set of alternatives through a scale formed by linguistic terms. In April 2017, an experimental analysis of this voting system was carried out during the French presidential elections, asking participants to assign a linguistic term from the above ordered qualitative scale to each presidential candidate (see Balinski and Laraki [7]).

Despite the extensive literature on the optimal number of response terms or categories that should be included in ordered qualitative scales, no clear conclusion has been reached (see Lissitz and Green [69], Cox [25], Oaster [82], Preston and Colman [86] and Matas [78], among others). In general, empirical evidence has shown that the use of very few response categories should be avoided, since it may harm the precision and validity of the scales if individuals are unable to correctly express their opinions when they answer questionnaires. On the other hand, due to the limited capacity of individuals to process and retain information, it is also undesirable to introduce a large number of response categories. For these reasons, most ordered qualitative scales usually employ between four and seven response categories (see Miller [80] and Lozano et al. [73]).

Regarding the design of surveys and questionnaires, the decision about whether to use an even or an odd number of response categories has also been discussed (see Garland [48], Johns [62], Baka et al. [5] and Díaz de Rada [29], among others). Ordered qualitative scales formed by an odd number of response categories are

generally used. These scales are usually balanced, that is, they include the same number of favorable and unfavorable terms separated by a central term that allows individuals to take a neutral position, e.g. {“strongly disagree”, “disagree”, “neither agree nor disagree”, “agree”, “strongly agree”}. However, when we want individuals to adopt a clear stance on a particular issue, we can use ordered qualitative scales formed by an even number of response categories: {“nada”, “poco”, “bastante”, “mucho”}. This 4-term ordered qualitative scale is quite often used in Barometers by the Spanish Center for Sociological Research (CIS, in Spanish).

Likert-type scales [68] are used to measure attitudes, knowledge, opinions and other qualitative aspects. This is an additive type scale that indicates an individual’s level of agreement or disagreement with a group of elements, statements or items. In general, these types of scale are balanced and include a neutral response category (see Lalla et al. [66] and Matas [78], among others). For instance, in the research carried out by Caballero Rodríguez [19], the following 5-point Likert scale was used: {1=“muy bajo”, 2=“bajo”, 3=“medio”, 4=“alto”, 5=“muy alto”} to determine the teaching staff’s level of satisfaction in relation to eight statements about entry and promotion prospects in the university.

As can be seen in the previous example, one characteristic of Likert-type scales is that they use the same ordered qualitative scale to assess all elements or statements. These elements are usually assigned the same weight or importance within the scale. Finally, once every element has been answered, the sum or average of the answers is calculated in order to determine the individual’s position on the variable being studied.

Ordered qualitative scales formed by linguistic terms are one of the most commonly used tools in qualitative research. However, some of the advantages of these scales are offset by the delicate statistical and mathematical analysis required, since there are no reliable established criteria about which techniques or procedures should be applied. Among the procedures proposed to overcome this problem we can highlight: the assignment of numerical values or scores to each of the linguistic terms on the scales (see Yates [110], Calatrava Requena [20], Roberts [88] and Fielding [34], among others), the use of fuzzy logic tools such as models based on fuzzy numbers and membership functions (see Lalla et al. [66] and Gil and González-Rodríguez [49], among others), and linguistic representation models such as *2-tuple model* introduced by Herrera and Martínez (see Herrera and Martínez [56] and Martínez et

al. [75], among others). This model has been applied in contexts involving linguistic information coming from multiple qualitative scales (see Herrera et al. [53]), as well as in contexts where unbalanced qualitative scales are used, that is, scales in which the linguistic terms are not symmetrically distributed around a central term (see Herrera et al. [54]).

The *2-tuple model* is based on the concept of symbolic translation to represent linguistic information. In this model, the result obtained after aggregating the indexes of the linguistic terms is represented by means of a pair of values called the linguistic 2-tuple. The tuple's first component represents the indexes of the linguistic terms, while the second indicates the distance between the result obtained and the closest linguistic term. Therefore, this procedure is mathematically equivalent to working with numerical values, once an integer has been associated to each linguistic term (see García-Lapresta [43]).

Other proposals made to handle qualitative data are: Ridit analysis in which the scores or “ridits” are obtained taking into account the relative frequency of the response categories (see Bross [16]), and conditional mean scoring functions that suggest a scoring function based on the assumption that the data are generated by underlying distributions (see Fielding [34] and Chen and Wang [22]).

Assigning numerical values or scores to the linguistic terms of the scales is the most popular method of analyzing information coming from ordered qualitative scales. However, the main criticism that can be made of this method is the subjectivity it introduces when numerical values are used to represent qualitative information (see Calatrava Requena [20]).

On the other hand, the use of numerical values may imply that there is some sort of proportional relationship between the response categories. However, this relationship is not necessarily present, since in ordinal scales the differences between the numerical values do not measure any magnitude that makes sense (see Grande Esteban [50] and Agresti [3], among others). Thus, for instance, if we assign equidistant scores to the linguistic terms of the scale used by the CIS: {“nada”, “poco”, “bastante”, “mucho”}, it could be considered that the linguistic terms are equally spaced, even though we have no evidence that the qualitative difference between “mucho” and “bastante” can be represented by the same numerical value as the difference between “poco” and “bastante” (see Calatrava Requena [20], Silva Aycaguer [96], Fleiss et al. [36], López-González [72], Chen and Wang [22] and Martínez

Martín [77], among others).

It is essential to bear in mind that, since the numerical values in ordered qualitative scales do not represent a quantitative relationship between the different response categories, arithmetic operations should not be performed with these numbers. In this sense Stevens [97, p. 679] point out:

The ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of data. On the other hand, for this “illegal” statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results. While the outlawing of this procedure would probably serve no good purpose, it is proper to point out that means and standard deviations computed on an ordinal scale are in error to the extent that the successive intervals on the scale are unequal in size.

Additionally, other authors such as Gardner [47], Roberts [88], Marcus-Roberts and Roberts [74], Knapp [64] and Franceschini [37], among others, have also widely questioned the widespread use of statistical techniques and procedures for the processing and analysis of qualitative data.

Despite the problems arising from the transformation of qualitative information into quantitative information, many national and international organizations continue to apply these techniques. This unadvisable praxis in the handling of ordered qualitative scales can sometimes lead to a distorted interpretation of the information, since it is possible to obtain different results depending on the codification of the response categories (see Stevens [97, 98], Merbitz et al. [79], Saaty [94], Franceschini [37], Franceschini et al. [38] and Gadrich et al. [41], among others). Likewise, Siegel [95, p. 3] highlights some drawbacks of using numerical scores in parametric statistical tests:

When these arithmetic processes are used on scores which are not truly numerical, they naturally introduce distortions in those data and thus throw in doubt any conclusions from the test.

In order to overcome these drawbacks and analyze ordinal data in an appropriate way, several procedures have been proposed in the literature based on dominance

criteria and the distribution functions of the variables. Blair and Lacy [12] propose a dispersion measure based on cumulative relative frequencies. In the dissertation of Martínez Martín [77], two generalizations of dispersion measures for ordinal variables are proposed, based on the frequencies of the response categories weighted by a discrepancy measure. Franceschini et al. [39] propose a new quality control method that uses information from ordinal scales. This procedure is based on classification and dispersion dominance criteria (see Franceschini and Romano [40] and Franceschini et al. [38]), and allows a statistical analysis to be performed without using traditional operators such as the sample mean or variance or an arbitrary numerical codification of the response categories on the scales. Bashkansky and Gadrich [10] also address quality processes through a method called *Proportion Ratio and Dispersion* (PR&D), which generates quality scales taking into account the relative frequencies of the categories and their dispersion. On the other hand, Herrero and Villar [58] handle categorical data through a procedure based on the distribution of the data and on dominance relations in pair-wise comparisons.

Another factor that can influence the analysis and handling of qualitative data is the way in which individuals perceive these scales. Usually, ordered qualitative scales are considered to be uniform, that is, there is the same proximity between all consecutive linguistic terms of the scale, which are symmetrically distributed around a central term or response category, for example: {“very bad”, “bad”, “fair”, “good”, “very good”}. However, common sense, the semantic value of the words, cultural differences or even the individual’s personality can at times cause ordered qualitative scales to be perceived as non-uniform (see Silva Aycaguer [96], Clarke [23], Javaras and Ripley [61], Hofmans et al. [59] and Martínez García and Martínez Caro [76], among others). An ordered qualitative scale can be considered to be non-uniform when individuals can perceive different psychological proximities between the consecutive linguistic terms of the scale. For instance, the ordered qualitative scale used by the CIS can be considered to be non-uniform if individuals appreciate that the linguistic term “bastante” is closer to “mucho” than to “poco”, or if the term “poco” is closer to “nada” than to “bastante”.

In the context of non-uniform ordered qualitative scales, it is not appropriated to assign numerical values or scores to the linguistic terms, since these may not properly reflect the individual’s perceptions of the proximities between the terms of the scales.

In this doctoral thesis, to handle non-uniform ordered qualitative scales in a purely ordinal way, we use the concept of the *ordinal proximity measure* introduced by García-Lapresta and Pérez-Román [45]. Ordinal proximity measures collect the information about how individuals perceive the proximities between the terms of scales using a set of ordinal degrees of proximity. It is important to note that these degrees of proximity are not numbers. They are abstract objects that represent different proximities.

The main objective of this thesis is the analysis, extension and application of ordinal proximity measures to different contexts. To do that, several procedures for generating and aggregating ordinal proximity measures have been designed, and new multi-criteria decision-making models that use these measures have been proposed.

In **Chapter 1**, the procedures for generating and aggregating ordinal proximity measures are analyzed. The concept of metrizable ordinal proximity measures is also introduced. These are ordinal proximity measures where the ordinal comparisons between the linguistic terms can be represented through a linear metric on the ordered qualitative scale.

Likewise, this chapter presents an algorithm that avoids problems and inconsistencies in the generation of metrizable ordinal proximity measures on ordered qualitative scales formed by four linguistic terms. This algorithm allows us to obtain the metrizable ordinal proximity measure for each expert through an appropriate sequence of questions about the proximities between the terms of the scale. Since experts can perceive different psychological proximities between the terms of an ordered qualitative scale, it is possible to obtain different metrizable ordinal proximity measures when applying the algorithm. In these cases, an aggregation problem of the experts' opinions is emerged within Social Choice Theory (see List [71]).

From the perspective of the Judgment Aggregation Theory (see Dietrich and List [30], List [70] and Grossi and Pigozzi [51], among others), several procedures that seek to obtain a collective judgment reflecting the opinions of a group of experts as faithfully as possible have been analyzed and proposed. Nevertheless, it is important to note that the choice of aggregation procedure is not trivial, since applying different voting systems may lead to different results and even inconsistencies.

In order to avoid problems and inconsistencies when experts' opinions are aggregated, in this chapter we propose an aggregation procedure based on weighted metrics. This procedure considers that the collective metrizable ordinal proximity

measure is that which minimizes the sum of distances (or squared distances) between itself and metrizable ordinal proximity measures of the experts.

This chapter also includes some examples to illustrate how the algorithm and the aggregation procedure work.

Chapter 2 presents an application of the voting system introduced by García-Lapresta and Pérez-Román [46] to data taken from the CIS Barometer.

This chapter includes a discussion of the methodology for grouping several response categories used by the CIS and other national and international organizations such as the Pew Research Center. This practice consists of grouping the results of certain response categories (for example “mucho” and “bastante” or “very good” and “somewhat good”), ignoring the rest of the response categories. The application of this methodology is implicitly assigning the same numerical value to the grouped response categories and a null value to the omitted categories, and therefore leads to a significant loss of qualitative information from these scales.

In order to avoid this loss of information and to consider all the response categories on the qualitative scales, we have ranked a set of alternatives taken from the CIS Barometer using the voting system proposed by García-Lapresta and Pérez-Román [46]. It is an ordinal procedure where the alternatives are ranked taking into account the pair of medians of ordinal degrees of proximity. For each of the alternatives considered, these ordinal degrees of proximity are obtained from comparing the highest linguistic term of the scale and the assessments given by the agents.

The *Majority Judgment* voting system (see Balinski and Laraki [8]) considers the median as a representative element of the collective assessment. In particular, this system takes the median as the collective grade when the number of assessments is odd, and the lower median when the number of assessments is even. However, there are some key differences between the *Majority Judgment* voting system and the procedure applied in this chapter. The voting system introduced by García-Lapresta and Pérez-Román [46] takes into account how individuals perceive the proximities between the terms of the scales, an aspect not considered in the *Majority Judgment* voting system. On the other hand, when the number of assessments is even, it can cause a significant loss of information if we only consider the lower median as collective assessment. For that reason, and in order to avoid a loss of information, the system applied in this chapter considers the two medians of ordinal degrees of

proximity.

The voting system proposed by García-Lapresta and Pérez-Román [46] also introduces a sequential tie-breaking procedure based on Balinski and Laraki [6, 8]. This consists of removing the pair of medians of degrees of proximity of the alternatives that are tied and selecting a new pair from among the remaining ordinal degrees of proximity. The tie-breaking procedure is repeated as many times as necessary until the ties are broken and the alternatives are ranked.

In **Chapter 3**, an ordinal multi-criteria decision-making procedure is introduced. This system is an extension of the procedure proposed by García-Lapresta and González del Pozo [44] and of the voting system introduced by García-Lapresta and Pérez-Román [46].

In particular, the proposed procedure ranks a set of alternatives assessed by a group of agents in relation to multiple criteria. The agents may use a non-uniform ordered qualitative scale to assess the alternatives.

Since the number of response categories of the scales is limited, agents may have doubts about which linguistic term to assign to each alternative. One way to incorporate imprecision in decision-making processes is to allow agents to assign several consecutive linguistic terms. In this regard, there are several papers in the literature in which such possibility is considered (see Torra [102], Agell et al. [1], Rodríguez et al. [89], Agell et al. [2], Falcó et al. [33], Rodríguez et al. [90] and Wang et al. [108], among others).

The procedure introduced in this chapter also incorporates the possibility of doubt, allowing agents to assess the alternatives through one or two consecutive linguistic terms when they are unsure about which linguistic term to use. To do that, we introduce an extension of ordinal proximity measures based on the *absolute order of magnitude spaces* introduced by Travé-Massuyès and Piera [103], and in the extensions of Falcó et al. [33] and Roselló et al. [91], among others.

In order to represent the weight of each criterion in the procedure, we propose to replicate the qualitative assessments of the alternatives as many times as necessary until these replications reflect the weightings of the criteria. Balinski and Laraki [8] also use replications in their proposal to extend the *Majority Judgment* voting system to multi-criteria problems. However, it is important to mention again that the *Majority Judgment* voting system does not consider whether qualitative scales are uniform or not and, in addition, it does not allow agents to use more than one

linguistic term in their assessments.

Finally, this chapter includes an application of the proposed procedure to rank a set of wines taking into account the most important sensory aspects in wine tasting: appearance, smell and taste, as well as the overall impression of the wines.

In the voting system and in the ordinal procedure considered in the previous chapters, the same ordered qualitative scale is always used to assess the alternatives. However, at times agents are required to express their assessments by means of several ordered qualitative scales formed by a different number of linguistic terms (see Herrera et al. [53], Herrera et al. [57] and de Andrés et al. [4], among others).

Chapter 4 proposes a multi-criteria decision-making procedure that manages information from different ordered qualitative scales. This procedure postpones for as long as possible the use of numerical values in order to preserve and respect the ordinal information of the scales. The proposed procedure allows agents to evaluate each criterion through a specific ordered qualitative scale. These ordered qualitative scales can be considered to be non-uniform and to be formed of a different number of linguistic terms.

Following the approaches of the Copeland rule and the multi-criteria decision method *PROMETHEE II* (see Copeland [24], Brans [14], Bouyssou [13] and Brans and De Smet [15]), the proposed procedure yields a ranking of alternatives based on the intuitive principle of compensation between advantages and disadvantages of each alternative in comparison with its opponents.

To show how the multi-criteria procedure works, we present an illustrative example in which a stochastic analysis is introduced to assess the robustness of the results obtained (see Lahdelma et al. [65], Tervonen and Lahdelma [101], Tervonen and Figueira [100] and Dias and Vetschera [28], among others). Likewise, this chapter also includes a comparison of the proposed procedure with the methods *SMAA-O* (see Lahdelma [65]) and *ZAPROS III* (see Larichev [67]).

Lastly, the dissertation finishes with a chapter of **conclusions and further research** that summarizes the main contributions of this doctoral thesis and presents some possible further research lines.

In addition, along the thesis writing process, the proposals and results obtained have been presented in different conferences and scientific meetings, as well as in several research seminars. They are detailed below:

National conferences and scientific meetings:

- ESTYLF 2018. XIX Congreso Español sobre Tecnologías y Lógica Fuzzy. Granada, 2018.
- SEIO 2018. XXXVII Congreso Nacional de Estadística e Investigación Operativa y XI Jornadas de Estadística Pública. Oviedo, 2018.
- X Encuentro Interuniversitario de Investigadores en Decisiones Públicas. Pamplona, 2017.
- V Jornadas de Trabajo sobre Sistemas de Votación. Valladolid, 2017.
- ESTYLF 2016. XVIII Congreso Español sobre Tecnologías y Lógica Fuzzy. San Sebastián, 2016.

International conferences and scientific meetings:

- DSRS-Turing 2019. 1st International Alan Turing Conference on Decision Support and Recommender Systems. London (United Kingdom), 2019.
- EURO 2019. 30th European Conference on Operational Research. Dublin (Ireland), 2019.
- ASMDA 2019. 18th Applied Stochastic Models and Data Analysis International Conference. Florence (Italy), 2019.
- EMPG 2018. Meeting of the European Mathematical Psychology Group. Genoa (Italy), 2018.
- EWG/MCDA 2018. 87th Meeting of the European Working Group on Multi-criteria Decision Aiding. Delft (The Netherlands), 2018.
- MOPGP 2017. 12th International Conference on Multiple Objective Programming and Goal Programming. Metz (France), 2017.
- III International Workshop on Proximity Data, Multivariate Analysis and Classification. Valladolid (Spain), 2017.

- EUSFLAT 2017. The 10th Conference of the European Society for Fuzzy Logic and Technology. Warsaw (Poland), 2017.
- IFORS 2017. 21st Conference of the International Federation of Operational Research Societies. Quebec City (Canada), 2017.
- MCDM 2017. 24th International Conference on Multiple Criteria Decision Making. Ottawa (Canada), 2017.
- Workshop on Future Directions in Computational Social Choice. Budapest (Hungary), 2016.

Research seminars:

- Research seminar organized by the Centro de Estudos de Gestão do Instituto Superior Técnico (CEG-IST). Instituto Superior Técnico (IST). Universidade de Lisboa. Lisbon (Portugal), 2019.
- Research seminar organized by the Centre for Business and Economics Research (CeBER). Faculdade de Economia. Universidade de Coimbra. Coimbra (Portugal), 2018.
- Research seminar organized by the Grupo de Investigación de Ciencias Sociales Aplicadas. Facultad de Ciencias Económicas y Empresariales. Universidad de Valladolid. Valladolid (Spain), 2018.
- Research seminar organized by the PhD Programme in Business Economics in collaboration with the PhD Programme in Economics. Facultad de Ciencias Económicas y Empresariales. Universidad de Valladolid. Valladolid (Spain), 2018.

Chapter 1

Metrizable ordinal proximity measures and their aggregation

Ordered qualitative scales formed by linguistic terms are frequently used for evaluating sets of alternatives in different decision-making problems. These scales are usually implicitly considered as uniform in the sense that the psychological proximity between consecutive terms is perceived as identical. However, sometimes agents can perceive different proximities between the linguistic terms of the scale, and an appropriate method is required for aggregating these perceptions. In this paper we introduce the notion of metrizable ordinal proximity measure, discuss some aggregation procedures and propose a method based on metrics for aggregating experts' opinions on proximities between linguistic terms on ordered qualitative scales.

[This chapter has been previously published (jointly with José Luis García-Lapresta and David Pérez Román) in *Information Sciences*, vol. 448-449, pp. 149-163, 2018; doi:10.1016/j.ins.2018.03.034.]

Links:

<https://www.sciencedirect.com/science/article/pii/S0020025518302081>

<http://uvadoc.uva.es/handle/10324/29339>

Chapter 2

Tratamiento ordinal de las escalas cualitativas utilizadas por el Centro de Investigaciones Sociológicas

Ordinal treatment of qualitative scales used by the Center for Sociological Research

Las escalas cualitativas formadas por términos lingüísticos son utilizadas por diversas disciplinas para determinar preferencias y diversos aspectos de la vida de los individuos. Aunque es habitual asignar números a las categorías de respuesta de estas escalas, no es adecuado utilizar codificaciones numéricas cuando los individuos perciben proximidades psicológicas distintas entre las categorías consecutivas de la escala, es decir, cuando las escalas cualitativas no son uniformes. En este trabajo se propone un procedimiento ordinal para jerarquizar un conjunto de alternativas a partir de las valoraciones mostradas por un grupo de individuos mediante una escala cualitativa no necesariamente uniforme. Dicho procedimiento está basado en las proximidades ordinales entre las categorías de respuesta de las escalas. El pro-

cedimiento propuesto se ilustra con un ejemplo tomado del Barómetro del Centro de Investigaciones Sociológicas (CIS) de mayo de 2011.

Qualitative scales formed by linguistic terms are used by different disciplines to determine preferences and different aspects of individuals' lives. Although it is usual to assign numbers to the response categories of scales, it is not suitable when individuals perceive different proximities between the consecutive categories of the scale, that is, when scales are not uniform. In this paper, an ordinal procedure is proposed to order a set of alternatives from the assessments given by a group of individuals through a qualitative scale not necessarily uniform. This procedure is based on ordinal proximities between the response categories of scales. The proposed procedure is illustrated with an example taken from the Barometer of the Center for Sociological Research of May 2011.

[This chapter has been previously published (jointly with José Luis García-Lapresta in *Revista de Métodos Cuantitativos para la Economía y la Empresa* 28, pp. 124-142, 2019.)

Links:

<https://www.upo.es/revistas/index.php/RevMetCuant/article/view/3788>

<http://uvadoc.uva.es/handle/10324/40517>]

Chapter 3

An ordinal multi-criteria decision-making procedure under imprecise linguistic assessments

Many decision-making problems such as quality control analysis, market surveys or sensory analysis require ordered qualitative scales, rather than numerical ones. It is very common to assign some cardinal mathematical objects, such as numerical values, intervals of real numbers or fuzzy numbers, to the linguistic terms of ordered qualitative scales. However, when agents perceive that the psychological proximity between each pair of consecutive terms of the scale is not identical, these conversions are meaningless and an ordinal approach to deal with these non-uniform ordered qualitative scales is more appropriate. The aim of this paper is to introduce an ordinal multi-criteria decision-making procedure for ranking alternatives in the setting of ordered qualitative scales that are nonnecessarily uniform. The possibility of doubt is also considered, by allowing agents to assign two consecutive terms of the scale when they hesitate. The proposed procedure is applied to a real case study in which nine experts assessed eight wines regarding different criteria.

[This chapter has been previously published (jointly with José Luis García-Lapresta in *European Journal of Operational Research* 279, pp. 159-167, 2019; doi:10.1016/j.ejor.2019.05.015.)

Links:

<https://www.sciencedirect.com/science/article/abs/pii/S0377221719304199>

<http://uvadoc.uva.es/handle/10324/36090>]

Chapter 4

Using different qualitative scales in a multi-criteria decision-making procedure

Many decision problems manage linguistic information assessed through several ordered qualitative scales. In these contexts, the main problem arising is how to aggregate this qualitative information. In this paper, we present a multi-criteria decision-making procedure that ranks a set of alternatives assessed by means of a specific ordered qualitative scale for each criterion. These ordered qualitative scales can be non-uniform and be formed by a different number of linguistic terms. The proposed procedure follows an ordinal approach by means of the notion of ordinal proximity measure that assigns an ordinal degree of proximity to each pair of linguistic terms of the qualitative scales. To manage the ordinal degree of proximity from different ordered qualitative scales, we provide a homogenization process. We also introduce a stochastic approach to assess the robustness of the conclusions.

[This chapter has been previously published (jointly with Luís C. Dias and José Luis García-Lapresta in *Mathematics* 2020, 8, 458; doi:10.3390/math8030458.)

Links:

<https://www.mdpi.com/2227-7390/8/3/458/htm>

<http://uvadoc.uva.es/handle/10324/40693>

Conclusiones y trabajo futuro

Finalmente, se presentan los principales resultados y contribuciones de esta tesis doctoral, así como algunas futuras líneas de investigación a desarrollar.

Conclusiones

En esta tesis doctoral se han analizado las escalas cualitativas no uniformes y se han presentado nuevas herramientas y modelos para el tratamiento ordinal de este tipos de escalas. Las principales contribuciones que se han alcanzado a lo largo del desarrollo de esta tesis doctoral se exponen a continuación.

En el **capítulo 1** se han estudiado diferentes procedimientos para generar y agregar medidas de proximidad ordinal. Algunas de las principales aportaciones de este capítulo son:

- La introducción del concepto de métrica lineal y de la noción de medida de proximidad ordinal metrizable, es decir, aquellas medidas de proximidad ordinal que se comportan como si las comparaciones entre los términos lingüísticos, se realizaran considerando la existencia de una métrica lineal sobre la escala cualitativa.
- El diseño de un algoritmo que permite obtener la medida de proximidad ordinal metrizable para cada experto, a través de una secuencia apropiada de preguntas sobre las proximidades entre los términos lingüísticos de la escala. Este algoritmo evita problemas e inconsistencias a la hora de generar medidas de proximidad ordinal metrizable en escalas cualitativas formadas por cuatro términos lingüísticos.
- Un análisis de las posibles inconsistencias que pueden surgir al aplicar diversos sistemas de votación para generar medidas de proximidad ordinal metrizable.

- La propuesta de un procedimiento de agregación que evita problemas e inconsistencias al determinar la medida de proximidad ordinal metrizable colectiva. Dicha medida de proximidad ordinal se obtiene minimizando la suma de distancias (o suma de distancias al cuadrado) entre sí misma y las medidas de proximidad ordinal metrizable de los expertos.

En el **capítulo 2** se ha abordado el tratamiento de las escalas cualitativas por parte de distintos organismos nacionales e internacionales. En particular, este capítulo se centra en discutir y analizar la metodología de agrupación de varias categorías de respuesta realizada por el CIS. Las principales conclusiones que pueden extraerse de este capítulo son las siguientes:

- La escala cualitativa de cuatro categorías de respuesta utilizada por el CIS: {“nada”, “poco”, “bastante”, “mucho”} puede ser percibida como no uniforme.
- La metodología de agrupación de varias categorías de respuesta utilizada por el CIS y por otros organismos nacionales e internacionales, supone una pérdida importante de la información cualitativa de las mismas. Esta práctica está asignando de forma implícita un mismo valor numérico a las categorías de respuesta agrupadas y valores nulos a las categorías omitidas.
- La importancia de tratar las escalas cualitativas de forma ordinal, ya que cuando las escalas cualitativas son consideradas como no uniformes, las codificaciones numéricas asignadas a los términos lingüísticos pueden no reflejar adecuadamente la manera en la que los agentes perciben las proximidades entre estos términos, dificultando así la representatividad e interpretación de los resultados obtenidos.

En el **capítulo 3** se ha presentado un procedimiento de toma de decisiones multi-criterio donde los agentes pueden valorar las alternativas a través de uno o dos términos lingüísticos consecutivos. Entre los principales resultados de este capítulo destacan:

- La extensión del sistema de votación aplicado en el capítulo 2 a un contexto multi-criterio. El procedimiento introducido en este capítulo réplica las valoraciones cualitativas dadas por los agentes a las alternativas para representar el peso de cada criterio.

- La introducción de la posibilidad de duda. El procedimiento permite a los agentes valorar las alternativas mediante la asignación de uno o dos términos lingüísticos consecutivos. Para incorporar la duda en el procedimiento, se ha planteado una extensión del concepto de medida de proximidad ordinal.
- La posibilidad de aplicar el procedimiento introducido en este capítulo a escenarios multi-criterio donde se permita a los agentes asignar dos términos lingüísticos consecutivos en caso de duda: sesiones de prueba o degustación dirigidas al lanzamiento de nuevos productos al mercado, procedimientos de control de calidad, análisis sensorial, etc.

Finalmente, en el **capítulo 4** se ha introducido un procedimiento multi-criterio para ordenar alternativas valoradas por un grupo de agentes a través de distintas escalas cualitativas. Entre las principales aportaciones cabe mencionar:

- El procedimiento propuesto maneja información procedente de diversas escalas cualitativas posponiendo todo lo posible el uso de valores numéricos. En el procedimiento, las escalas cualitativas empleadas para valorar los criterios pueden ser no uniformes e incluso estar formadas por un número distinto de términos lingüísticos. Esto supone una diferencia importante con respecto a otros procedimientos multi-criterio, los cuales suelen emplear la misma escala cualitativa para valorar todos los criterios.
- La introducción de un proceso de homogeneización para agregar los grados de proximidad ordinal procedentes de las diferentes escalas cualitativas.
- La aplicabilidad del procedimiento propuesto a problemas reales: diagnóstico clínico, encuestas de satisfacción y calidad de vida, procesos de control de calidad, etc.

Trabajo futuro

Muchos procedimientos de toma de decisiones utilizan información procedente de una o varias escalas cualitativas para asignar ponderaciones y valorar alternativas. De igual modo, cada vez es más frecuente la consideración de factores cualitativos y cuantitativos en los procedimientos de toma de decisiones. Por consiguiente, es

necesario desarrollar procedimientos fiables que manejen y agreguen adecuadamente información heterogénea.

A continuación se proponen algunos posibles trabajos a realizar siguiendo la línea de investigación de esta tesis:

- Incorporar la posibilidad de duda al procedimiento multi-criterio propuesto en el capítulo 4.
- Aplicar los procedimientos propuestos en esta tesis a bases de datos del CIS con el fin de analizar el estado de la opinión pública en relación con los partidos políticos y sus dirigentes.
- Extender las medidas de proximidad ordinal a procedimientos de toma de decisiones multi-criterio en los que se manejen dos escalas cualitativas distintas: una escala cualitativa para determinar el peso de los expertos en el procedimiento según su experiencia; y otra escala cualitativa distinta para evaluar las alternativas.
- Desarrollar nuevos procedimientos de toma de decisiones que combinen información cualitativa y cuantitativa.

Conclusions and further research

Finally, we present the main findings and contributions of this dissertation, as well as some future research lines to develop.

Conclusions

In this doctoral thesis, we have analyzed non-uniform ordered qualitative scales and presented new tools and models for the ordinal treatment of this type of scale. The main contributions made through this dissertation are set out below.

In **Chapter 1** different procedures for generating and aggregating ordinal proximity measures have been examined. Some of the main contributions of this chapter are:

- The introduction of the linear metric concept and the notion of the metrizable ordinal proximity measure, that is, those ordinal proximity measures that behave as if the ordinal comparisons between the linguistic terms were made considering the existence of a linear metric on the ordered qualitative scale.
- The design of an algorithm that enables us to obtain the metrizable ordinal proximity measure for each expert, through an appropriate sequence of questions about the proximities between the linguistic terms of the scale. This algorithm avoids problems and inconsistencies in the generation of metrizable ordinal proximity measures on ordered qualitative scales formed by four linguistic terms.
- An analysis of the possible inconsistencies that may arise when different voting systems are applied to generate metrizable ordinal proximity measures.

- The proposal of an aggregation procedure that avoids problems and inconsistencies when the collective metrizable ordinal proximity measure is determined. This ordinal proximity measure is obtained by minimizing the sum of distances (or the sum of squared distances) between itself and the metrizable ordinal proximity measures of the experts.

Chapter 2 addresses the handling of ordered qualitative scales by different national and international organizations. In particular, this chapter focuses on discussing and analyzing the methodology used by the CIS, consisting of grouping together several response categories. The main conclusions that can be drawn from this chapter are:

- The 4-term ordered qualitative scale used by the CIS: {“nada”, “poco”, “bastante”, “mucho”}, it can be perceived as non-uniform.
- The methodology used by the CIS and other national and international organizations, consisting of grouping together several response categories, leads to a significant loss of qualitative information. This practice implicitly assigns the same numerical value to each of the grouped response categories and a null value to the omitted categories.
- The importance of treating ordered qualitative scales in an ordinal way, since when ordered qualitative scales are considered to be non-uniform, the numerical codifications assigned to the linguistic terms may not adequately reflect how agents perceive the proximities between the terms, thus hampering the representativeness and interpretation of the findings.

Chapter 3 presents a multi-criteria decision-making procedure where agents can assess alternatives by means of one or two consecutive linguistic terms. The main results of this chapter include:

- The extension of the voting system applied in Chapter 2 to a multi-criteria context. The procedure introduced in this chapter replicates the qualitative assessments given by the agents to the alternatives to represent the weight of each criterion.

- The introduction of the possibility of doubt. The procedure allows agents to assess the alternatives by means of one or two consecutive linguistic terms. To incorporate doubt into the procedure, an extension of the concept of the ordinal proximity measure has been provided.
- The possibility of applying the procedure introduced in this chapter to multi-criteria settings where agents are allowed to assign two consecutive linguistic terms in the event of doubt: tasting sessions aimed at launching new products on the market, quality control processes, sensory evaluations, etc.

Finally, **Chapter 4** introduces a multi-criteria decision-making procedure for ranking alternatives assessed by a group of agents through different ordered qualitative scales. The main contributions include:

- The proposed procedure manages information coming from several ordered qualitative scales, postponing as much as possible the use of numerical values. In the procedure, the ordered qualitative scales used for assessing the criteria can be non-uniform and can even be formed by a different number of linguistic terms. That introduces an important difference compared to other multi-criteria procedures which quite often use the same ordered qualitative scale to evaluate all criteria.
- The introduction of a homogenization process for aggregating the ordinal degrees of proximity coming from the different ordered qualitative scales.
- The applicability of the proposed procedure to real problems: clinical diagnosis, satisfaction and quality of life surveys, quality control processes, etc.

Further research

Many decision-making procedures use information from one or more ordered qualitative scales to assign weights and assess alternatives. Likewise, the consideration of qualitative and quantitative factors is increasingly common in decision-making procedures. Therefore, the development of reliable procedures for adequately processing and aggregating heterogeneous information is required.

Below are some proposals for future work to extend the area of research in this doctoral thesis:

- To incorporate the possibility of doubt into the multi-criteria procedure proposed in Chapter 4.
- To apply the procedures proposed in this dissertation to CIS datasets, in order to analyze public opinion on the political parties and their leaders.
- To extend ordinal proximity measures to multi-criteria decision-making procedures that require the handling of two different ordered qualitative scales: one scale for determining the weight of the experts in the procedure according to their expertise; and another different ordered qualitative scale for assessing the alternatives.
- To develop new decision-making procedures combining qualitative and quantitative information.

Bibliografía/Bibliography

- [1] Agell, N., Sánchez, M., Prats, F., Roselló, L.: Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions. *Information Sciences* 209, pp. 49-60, 2012.
- [2] Agell, N., Sánchez, G., Sánchez, M., Ruiz, F.J.: Selecting the best taste: a group decision-making application to chocolates design. *Proceedings of the 2013 IFSA-NAFIPS Joint Congress*, Edmonton, pp. 939-943, 2013.
- [3] Agresti, A.: *Analysis of Ordinal Categorical Data*. Second edition. John Wiley & Sons, 2010.
- [4] de Andrés, R., García-Lapresta, J.L., Martínez, L.: A multi-granular linguistic model for management decision-making in performance appraisal. *Soft Computing* 14, pp. 21-34, 2010.
- [5] Baka, A., Figgou, L., Triga, V.: ‘Neither agree, nor disagree’: a critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance* 5, pp. 244-263, 2012.
- [6] Balinski, M., Laraki, R.: A theory of measuring, electing and ranking. *Proceedings of the National Academy of Sciences of the United States of America* 104, pp. 8720-8725, 2007.
- [7] Balinski, M., Laraki, R.: Election by Majority Judgement: Experimental Evidence. *Ecole Polytechnique-Centre National de la Recherche Scientifique* 2007 (28), 2007.
- [8] Balinski, M., Laraki, R.: *Majority Judgment: Measuring, Ranking, and Electing*. The MIT Press, Cambridge, MA, 2011.

- [9] Balinski, M., Laraki, R.: Majority judgment vs. majority rule. *Social Choice and Welfare* 54, pp. 429-461, 2020.
- [10] Bashkansky, E., Gadrich, T.: Evaluating quality measured on a ternary ordinal scale. *Quality and Reliability Engineering International* 24, pp. 957-971, 2008.
- [11] Beyth-Marom, R.: How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting*, 1, pp. 257-269, 1982.
- [12] Blair, J., Lacy, M.G.: Statistics of ordinal variation. *Sociological Methods & Research* 28, pp. 251-280, 2000.
- [13] Bouyssou, D.: Ranking methods based on valued preference relations: A characterization of the net flow method. *European Journal of Operation Research* 60, pp. 61-67, 1992.
- [14] Brans, J.P.: L'ingénierie de la décision: Elaboration d'instruments d'aide à la décision. La méthode PROMETHEE. In: *L'aide à la Décision: Nature, Instruments et Perspectives D'Avenir* (eds. R. Nadeau, M. Landry). Presses de l'Université Laval, Quebec, 1982, pp. 183-213, 1982.
- [15] Brans, J.P., De Smet, Y.: PROMETHEE methods. In: *Multiple Criteria Decision Analysis: State of the Art Surveys* (eds. S. Greco, M. Ehrgott, J. Figueira). International Series in Operations Research & Management Science 233. Springer, pp. 187-219, 2016.
- [16] Bross, I.D.: How to use ridit analysis. *Biometrics* 14, pp. 18-38, 1958.
- [17] Budescu, D.V., Wallsten, T.S.: Subjective estimation of precise and vague uncertainties. In: *Judgmental Forecasting* (eds. G. Wright, P. Ayton). John Wiley & Sons, pp. 63-82, 1987.
- [18] Budescu, D.V., Weinberg, S., Wallsten, T.S.: Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance* 14, pp. 281-294, 1988.
- [19] Caballero Rodríguez, K.: Nivel de satisfacción del profesorado universitario hacia los sistemas de evaluación. *Revista de Educación* 360, pp. 483-508, 2013.

- [20] Calatrava Requena, J.: El método de “Ridit Analysis” para comparación múltiple de frecuencias muestrales ordenadas en niveles cualitativos: Aplicación al análisis de test y encuestas en Sociología. *Revista Española de la Opinión Pública* 47, pp. 123-135, 1977.
- [21] Chen, S.J., Hwang, C.L.: Fuzzy multiple attribute decision making: Methods. In: *Fuzzy Multiple Attribute Decision Making*. Lecture Notes in Economics and Mathematical Systems 375. Springer, pp. 289-486, 1992.
- [22] Chen, H.C., Wang, N.S.: The assignment of scores procedure for ordinal categorical data. *The Scientific World Journal*. Volume 2014, article ID 3042132014. <https://doi.org/10.1155/2014/304213>, 2014.
- [23] Clarke, I.: Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality* 15, pp. 137-152, 2000.
- [24] Copeland, A.H.: *A “Reasonable” Social Welfare Function*. Mimeo. University of Michigan Seminar on Applications of Mathematics to the Social Sciences. University of Michigan, USA, 1951.
- [25] Cox, E.P.: The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research* 17, pp. 407-422, 1980.
- [26] Delgado, M., Herrera, F., Herrera-Viedma, E., Martínez, L.: Combining numerical and linguistic information in group decision making. *Information Sciences* 107, pp. 177-194, 1998.
- [27] Dhami, M.K., Mandel, D.R.: Words or Numbers? Communicating Probability in Intelligence Analysis. *American Psychologist*. <https://doi.org/10.31234/osf.io/zkvcf>, 2020.
- [28] Dias, L.C., Vetschera, R.: On generating utility functions in Stochastic Multi-criteria Acceptability Analysis. *European Journal of Operational Research* 278, pp. 672-685, 2019.
- [29] Díaz de Rada, V.: Calidad de los datos de preguntas de batería en encuestas presenciales: una comparación de un estudio con cuestionario en papel y en formato electrónico. *Revista Española de Investigaciones Sociológicas* 152, pp. 167-178, 2015.

- [30] Dietrich, F., List, C.: Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics* 19, pp. 391-424, 2007.
- [31] Doumpos, M., Zopounidis, C.: *Multicriteria Decision Aid Classification Methods*. Springer, 2002.
- [32] Durán, A., Ocaña, A.C., Cañadas, I., Pérez, F.J.: Construcción de cuestionarios para encuestas: el problema de la familiaridad de las opciones de respuesta. *Metodología de Encuestas* 2, pp. 27-60, 2000.
- [33] Falcó, E., García-Lapresta, J.L., Roselló, L.: Allowing agents to be imprecise: A proposal using multiple linguistic terms. *Information Sciences* 258, pp. 249-265, 2014.
- [34] Fielding, A.: Scoring functions for ordered classifications in statistical analysis. *Quality & Quantity* 27 pp. 1-17, 1993.
- [35] Figueira, J.R., Greco, S., Roy, B., Slowinski, R.: ELECTRE Methods: Main Features and Recent Developments. In: *Handbook of Multicriteria Analysis* (eds. C. Zopounidis, P. Pardalos). Analysis. Applied Optimization 103. Springer, pp. 51-89, 2010.
- [36] Fleiss, J.L., Levin, B., Paik, M.C.: *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 2003.
- [37] Franceschini, F.: *Advanced Quality Function Deployment*. CRC Press, 2001.
- [38] Franceschini, F., Galetto, M., Varetto, M.: Qualitative ordinal scales: the concept of ordinal range. *Quality Engineering* 16, pp. 515-524, 2004.
- [39] Franceschini, F., Galetto, M., Varetto, M.: Ordered samples control charts for ordinal variables. *Quality and Reliability Engineering International* 21, pp. 177-195, 2005.
- [40] Franceschini, F., Romano, D.: Control chart for linguistic variables: a method based on the use of linguistic quantifiers. *International Journal of Production Research* 37, pp. 3791-3801, 1999.

- [41] Gadrich, T., Bashkansky, E., Zitikis, R.: Assessing variation: a unifying approach for all scales of measurement. *Quality & Quantity* 49, pp. 1145-1167, 2015.
- [42] Gambará, H.: Incertidumbre y probabilidad subjetiva en la teoría de decisión conductual. *Revista de Psicología General y Aplicada: Revista de la Federación Española de Asociaciones de Psicología* 44, pp. 199-208, 1991.
- [43] García-Lapresta, J.L.: Reciprocal and linguistic preferences. In: *On Fuzziness. A Homage to Lotfi A. Zadeh* (eds. R. Seising, E. Trillas, C. Moraga, S. Termini). *Studies in Fuzziness and Soft Computing* 298. Springer, pp. 193-197, 2013.
- [44] García-Lapresta, J.L., González del Pozo, R.: An ordinal multi-criteria decision-making procedure in the context of uniform qualitative scales. In: *Soft Computing Applications for Group Decision-making and Consensus Modeling* (eds. M. Collan, J. Kacprzyk). *Studies in Fuzziness and Soft Computing* 357. Springer, pp. 297-304, 2018.
- [45] García-Lapresta, J.L., Pérez-Román, D.: Ordinal proximity measures in the context of unbalanced qualitative scales and some applications to consensus and clustering. *Applied Soft Computing* 35, pp. 864-872, 2015.
- [46] García-Lapresta, J.L., Pérez-Román, D.: Aggregating opinions in non-uniform ordered qualitative scales. *Applied Soft Computing* 67, pp. 652-657, 2018.
- [47] Gardner, P.L.: Scales and statistics. *Review of Educational Research* 45 pp. 43-57, 1975.
- [48] Garland, R.: The mid-point on a rating scale: Is it desirable? *Marketing Bulletin* 2, pp. 66-70, 1991.
- [49] Gil, M.Á., González-Rodríguez, G.: Fuzzy vs. Likert scale in Statistics. In: *Combining Experimentation and Theory* (eds. E. Trillas, P. Bonissone, L. Magdalena, J. Kacprzyk). *Studies in Fuzziness and Soft Computing* 271. Springer, pp. 407-420, 2012.
- [50] Grande Esteban, I.: *Marketing de los Servicios*. ESIC Editorial, 2005.

-
- [51] Grossi, D., Pigozzi, G.: *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool, 2014.
- [52] Herrera, F., Alonso, S., Chiclana, F., Herrera-Viedma, E.: Computing with words in decision making: Foundations, trends and prospects. *Fuzzy Optimization and Decision Making* 8, pp. 337-364, 2009
- [53] Herrera, F., Herrera-Viedma, E., Martínez, L.: A fusion approach for managing multi-granularity linguistic term sets in decision making. *Fuzzy Sets and Systems* 114, pp. 43-58, 2000.
- [54] Herrera, F., Herrera-Viedma, E., Martínez, L.: A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems* 16, pp. 354-370, 2008.
- [55] Herrera, F., Herrera-Viedma, E., Verdegay, J.L.: A sequential selection process in group decision making with a linguistic assessment approach. *Information Sciences* 85, pp. 223-239, 1995
- [56] Herrera, F., Martínez, L.: A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems* 8, pp. 746-752, 2000.
- [57] Herrera, F., Martínez, L., Sánchez, P.J.: Managing non-homogeneous information in group decision making. *European Journal of Operational Research* 166, pp. 115-132, 2005.
- [58] Herrero, C., Villar, A.: On the comparison of group performance with categorical data. *PLoS ONE* 8 (12): e84784. Doi:10.1371/journal.pone.0084784, 2013.
- [59] Hofmans, J., Theuns, P., Van Acker, F.: Combining quality and quantity. A psychometric evaluation of the self-anchoring scale. *Quality & Quantity* 43, pp. 703-716, 2009.
- [60] Jaffe-Katz, A., Budescu, D.V., Wallsten, T.S.: Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition* 17, pp. 249-264, 1989.

- [61] Javaras, K.N., Ripley, B.D.: An “unfolding” latent variable model for Likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association* 102, pp. 454-463, 2007.
- [62] Johns, R.: One size doesn’t fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties* 15, pp. 237-264, 2005.
- [63] Kadzinski, M., Greco, S., Slowinski, R.: Selection of a representative value function in robust multiple criteria ranking and choice. *European Journal of Operational Research* 217, pp. 541-553, 2012.
- [64] Knapp, T.R.: Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research* 39, pp. 121-123, 1990.
- [65] Lahdelma, R., Miettinen, K., Salminen, P.: Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). *European Journal of Operational Research* 147, pp. 117-127, 2003.
- [66] Lalla, M., Facchinetti, G., Mastroleo, G.: Ordinal scales and fuzzy set systems to measure agreement: An application to the evaluation of teaching activity. *Quality & Quantity* 38, pp. 577-601, 2005.
- [67] Larichev, O.I.: Ranking multicriteria alternatives: The method ZAPROS III. *European Journal of Operational Research* 131, pp. 550–558, 2001.
- [68] Likert, R.: A technique for the measurement of attitudes. *Archives of Psychology* 22, pp. 1-55, 1932.
- [69] Lissitz, R.W., Green, S.B.: Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology* 60, pp. 10-13, 1975.
- [70] List, C.: The theory of judgment aggregation: An introductory review. *Synthese* 187, pp. 179-207, 2012.
- [71] List, C.: Social Choice Theory. In: *The Stanford Encyclopedia of Philosophy* (ed. E. N. Zalta). Winter 2013 Edition, 2013.
<https://plato.stanford.edu/archives/win2013/entries/social-choice>.
- [72] López-González, E.: Sugerencias para el análisis de escalas con métrica delicada. *Revista Iberoamericana de Evaluación Educativa* 5, pp. 84-105, 2012.

- [73] Lozano, L.M, García-Cueto, E., Muñiz, J.: Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 4, pp. 73-79, 2008.
- [74] Marcus-Roberts, H.M., Roberts, F.S.: Meaningless statistics. *Journal of Educational Statistics* 12, pp. 383-394, 1987.
- [75] Martínez, L., Rodríguez, R.M., Herrera, F.: *The 2-tuple Linguistic Model. Computing with Words in Decision Making*. Springer, 2015.
- [76] Martínez García, J.A., Martínez Caro, L.: La medición de la satisfacción del consumidor de servicios deportivos a través de la lógica borrosa. *Revista de Psicología del Deporte* 19, pp. 41-58, 2010.
- [77] Martínez Martín, N.: Medidas de dispersión ordinal: aplicaciones a la medición del impacto de etiquetas lingüísticas y a problemas de decisión multicriterio con operadores OWA. PhD Dissertation. Universidad Complutense de Madrid (Madrid), 2020.
- [78] Matas, A.: Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista Electrónica de Investigación Educativa* 20, pp. 38-47, 2018.
- [79] Merbitz, C., Morris, J., Grip, J.C.: Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation* 70, pp. 308-312, 1989.
- [80] Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63, pp. 81-97, 1956.
- [81] Novák, V.: Fuzzy set theory as a theory of vagueness. In: *Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory* (eds. J. Kacprzyk, M. Fedrizzi). Theory and Decision Library 18, Springer, pp. 28-48, 1990.
- [82] Oaster, T.R.F.: Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills* 68, pp. 549-550, 1989.
- [83] Pawlak, Z.: Rough sets. *International Journal of Computer and Information Sciences* 11, pp. 341-356, 1982.

- [84] Pawlak, Z.: Vagueness: A rough set view. In: *Structures in Logic and Computer Science* (eds. J. Mycielski, A. Salomaa). Lecture Notes in Computer Science 1261. Springer, pp. 106-117, 1997.
- [85] Pawlak, Z., Slowinski, R.: Rough set approach to multi-attribute decision analysis. *European Journal of Operational Research* 72, pp. 443-459, 1994.
- [86] Preston, C.C., Colman, A.M.: Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica* 104, pp. 1-15, 2000.
- [87] Prieto, P.: La escala gráfica difusa: una alternativa a la asignación de magnitud a las respuestas psicológicas. PhD Dissertation. Universidad de la Laguna (Tenerife), 1991.
- [88] Roberts, F.S.: *Measurement Theory*. Addison-Wesley Publishing Company, Reading, 1979.
- [89] Rodríguez, R.M., Martínez, L., Herrera, F.: Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, 20, pp. 109-119, 2012.
- [90] Rodríguez, R.M., Martínez, L., Torra, V., Xu, Z.S., Herrera, F.: Hesitant fuzzy sets: State of the art and future directions. *International Journal of Intelligent Systems* 29, pp. 495-524, 2014.
- [91] Roselló, L., Sánchez, M., Agell, N., Prats, F., Mazaira, F.A.: Using consensus and distances between generalized multi-attribute linguistic assessments for group decision-making. *Information Fusion* 17, pp. 83-92, 2014.
- [92] Roy, B.: *Méthodologie Multicritère d'Aide à la Décision*. Economica, Paris, 1985.
- [93] Roy, B.: *Multicriteria Methodology for Decision Aiding*. Springer, 1996.
- [94] Saaty, T.L.: How to make a decision: the analytic hierarchy process. *European Journal of Operational Research* 48, pp. 9-26, 1990.
- [95] Siegel, S.: *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 1956.

- [96] Silva Aycaguer, L.C.: *Cultura Estadística e Investigación Científica en el Campo de la Salud: una Mirada Crítica*. Ediciones Díaz de Santos, 1997.
- [97] Stevens, S.S.: On the theory of scales of measurement. *Science* 103, pp. 677-680, 1946.
- [98] Stevens, S.S.: *Mathematics, Measurement, and Psychophysics*. John Wiley & Sons, 1951.
- [99] Teigen, K.H.: The language of uncertainty. *Acta Psychologica* 68, pp. 27-38, 1988.
- [100] Tervonen, T., Figueira, J.R.: A survey on stochastic multicriteria acceptability analysis methods. *Journal of Multi-Criteria Decision Analysis* 15, pp. 1-14, 2008.
- [101] Tervonen, T., Lahdelma, R.: Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research* 178, pp. 500-513, 2007.
- [102] Torra, V.: Hesitant fuzzy sets. *International Journal of Intelligent Systems* 25, pp. 529-539, 2010.
- [103] Travé-Massuyès, L., Piera, N.: The orders of magnitude models as qualitative algebras. *Proceedings of the 11th International Joint Conference on Artificial Intelligence* 2, pp. 1261-1266, 1989.
- [104] Von Winterfeldt, D., Edwards, W.: *Decision Analysis and Behavioral Research*. Cambridge University Press. Cambridge, 1986.
- [105] Wallsten, T.S., Budescu, D.V., Rapoport, A., Zwick, R., Forsyth, B.: Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General* 115, pp. 348-365, 1986.
- [106] Wallsten, T.S., Budescu, D.V., Zwick, R.: Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science* 39, pp. 176-190, 1993.

- [107] Wallsten, T.S., Fillenbaum, S., Cox, J.A.: Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language* 25, pp. 571-587, 1986.
- [108] Wang, H., Xu, Z., Zeng, X.J.: Hesitant fuzzy linguistic term sets for linguistic decision making: Current developments, issues and challenges. *Information Fusion* 43, pp. 1-12, 2018.
- [109] Windschitl, P.D., Wells, G.L.: Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied* 2, pp. 343-364, 1996.
- [110] Yates, F.: The analysis of contingency tables with groupings based on quantitative characters. *Biometrika* 35, pp. 176-181, 1948.
- [111] Zadeh, L.A.: Fuzzy sets. *Information and Control* 8, pp. 338-353, 1965.
- [112] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part I. *Information Sciences* 8, pp. 199-249, 1975; Part II. *Information Sciences* 8, pp. 301-357, 1975; Part III. *Information Sciences* 9, pp. 43-80, 1975.
- [113] Zadeh, L.A.: Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems* 4, pp. 103-111, 1996.
- [114] Zadeh, L.A.: The key roles of information granulation and fuzzy logic in human reasoning, Concept formulation and computing with words. In: *Proceedings of IEEE 5th International Fuzzy Systems* vol. 1 p. 1, 1996.
- [115] Zimmer, A.C.: Verbal vs. numerical processing of subjective probabilities. In: *Decision Making under Uncertainty* (ed. R.W. Scholz). Elsevier, pp. 159-182, 1983.
- [116] Zimmer, A.C.: A model for the interpretation of verbal predictions. *International Journal of Man-Machine Studies* 20, pp. 121-134, 1984.