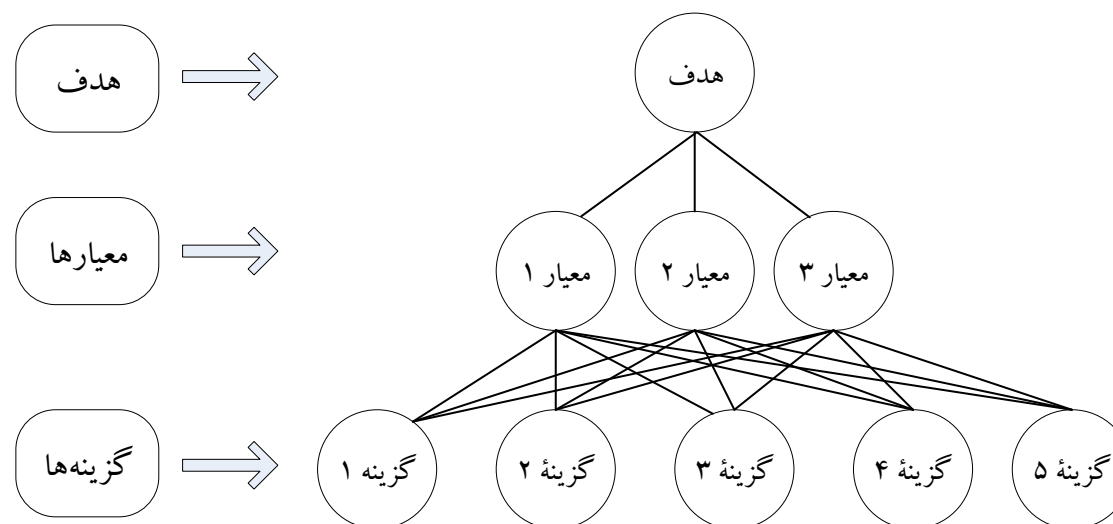


رتبه‌بندی حوزه‌های پژوهشی زبانشناسی رایانشی با روش تحلیل سلسله‌مراتبی (AHP)

۱. روش پژوهش

یکی از این ابزارهای پشتیبانی تصمیم‌گیری گروهی، فرایند تحلیل سلسله‌مراتبی است که می‌تواند تأثیر متقابل و مشارکت افراد در تصمیم‌گیری را افزایش دهد (قدسی‌پور ۱۳۷۹).

فرایند تحلیل سلسله‌مراتبی (AHP) یکی از روش‌های تصمیم‌گیری چندمعیاره است که ال‌ساعتی^۱ (۱۹۸۰) آن را مطرح کرد.



^۱. Saati, El.

مثال AHP



۱. تشکیل گروه‌های تصمیم‌ساز

۲. تعیین گزینه‌ها

۳. تعیین معیارها

منظور از گزینه‌ها یا کاندیداها همان موضوعات، اشیا و مواردی هستند که قصد انتخاب یکی از آنها را داریم. معیارها که خود می‌توانند شامل یک یا چند زیرمعیار باشند، ملاکهای تصمیم‌گیری هستند و هدف انتخاب گزینه نهایی است. به طور خلاصه در فرایند تحلیل سلسله‌مراتبی اولین قدم تشکیل گروه‌های تصمیم‌ساز برای تجزیه و تحلیل مسئله و در نهایت ارائه مدلی سلسله‌مراتبی به صورت بالا از مسئله مورد نظر است (قدسی‌پور ۱۳۷۹).

مقایسه زوجی معیارها و وزن‌دهی آنها

	پیشینه	سن	کاریزما	تحصیلات	بردار ویژه
پیشینه	1	7	3	4	0.547
سن	1/7	1	1/5	1/3	0.056
کاریزما	1/3	5	1	3	0.270
تحصیلات	1/4	3	1/3	1	0.127

مقیاس نه درجهٔ ساعتی

شرح	تعریف	درجه اهمیت
دو عنصر، اهمیت یکسانی دارند.	اهمیت یکسان	۱
یک عنصر، نسبت به عنصر دیگر، نسبتاً ترجیح داده می‌شود.	نسبتاً مرجح	۳
یک عنصر، نسبت به عنصر دیگر، زیاد ترجیح داده می‌شود.	ترجیح زیاد	۵
یک عنصر، نسبت به عنصر دیگر، بسیار زیاد ترجیح داده می‌شود.	ترجیح بسیار زیاد	۷
یک عنصر، نسبت به عنصر دیگر، فوق‌العاده زیاد ترجیح داده می‌شود.	ترجیح فوق‌العاده زیاد	۹
هنگامی که عنصر A با عنصر B مقایسه می‌شود، یکی از اعداد بالا به آن اختصاص می‌یابد. در مقایسه‌ی عنصر A با A، مقدار معکوس آن عدد اختصاص می‌یابد.		

مقایسه زوجی گزینه‌ها

پیشینه	مادلین	سوف	راجر
مادلین	1	1/4	4
سوف	4	1	9
راجر	1	1/9	1

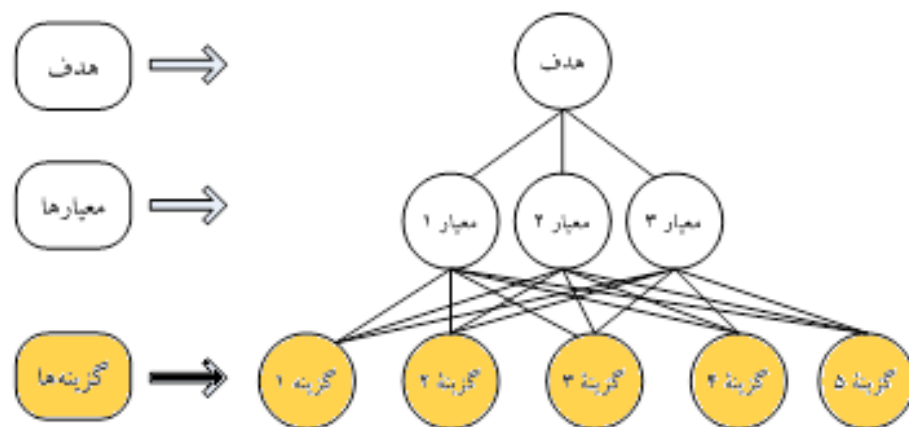
وزن گزینه‌ها بر اساس معیارها

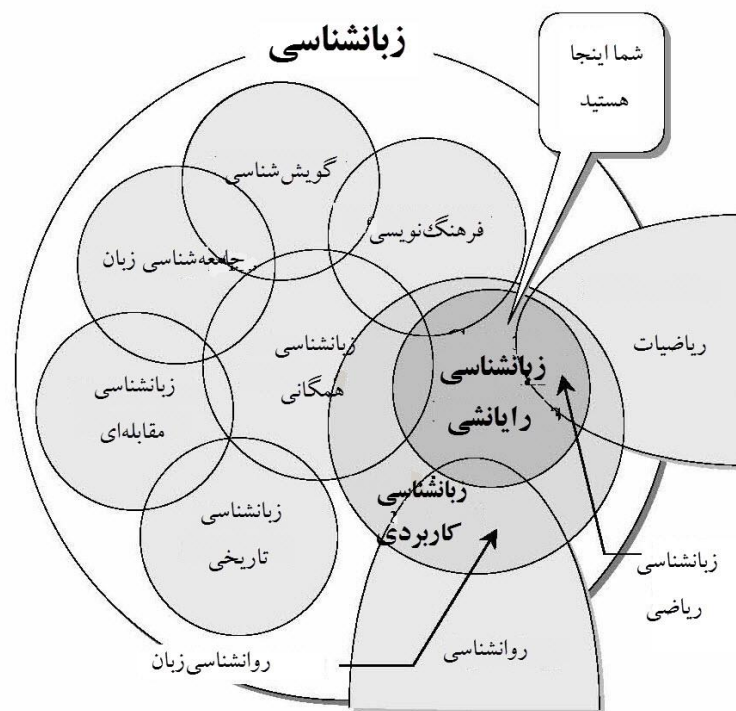
ماددین	پیشینه	سن	کارزما	تحصیلات
0.217	0.265	0.743	0.188	
0.717	0.672	0.194	0.081	
0.066	0.063	0.063	0.731	

۱-۱ هدف

اولویت‌بندی حوزه‌های پژوهشی در گروه زبان‌شناسی رایانشی

۱-۲ گزینه‌ها

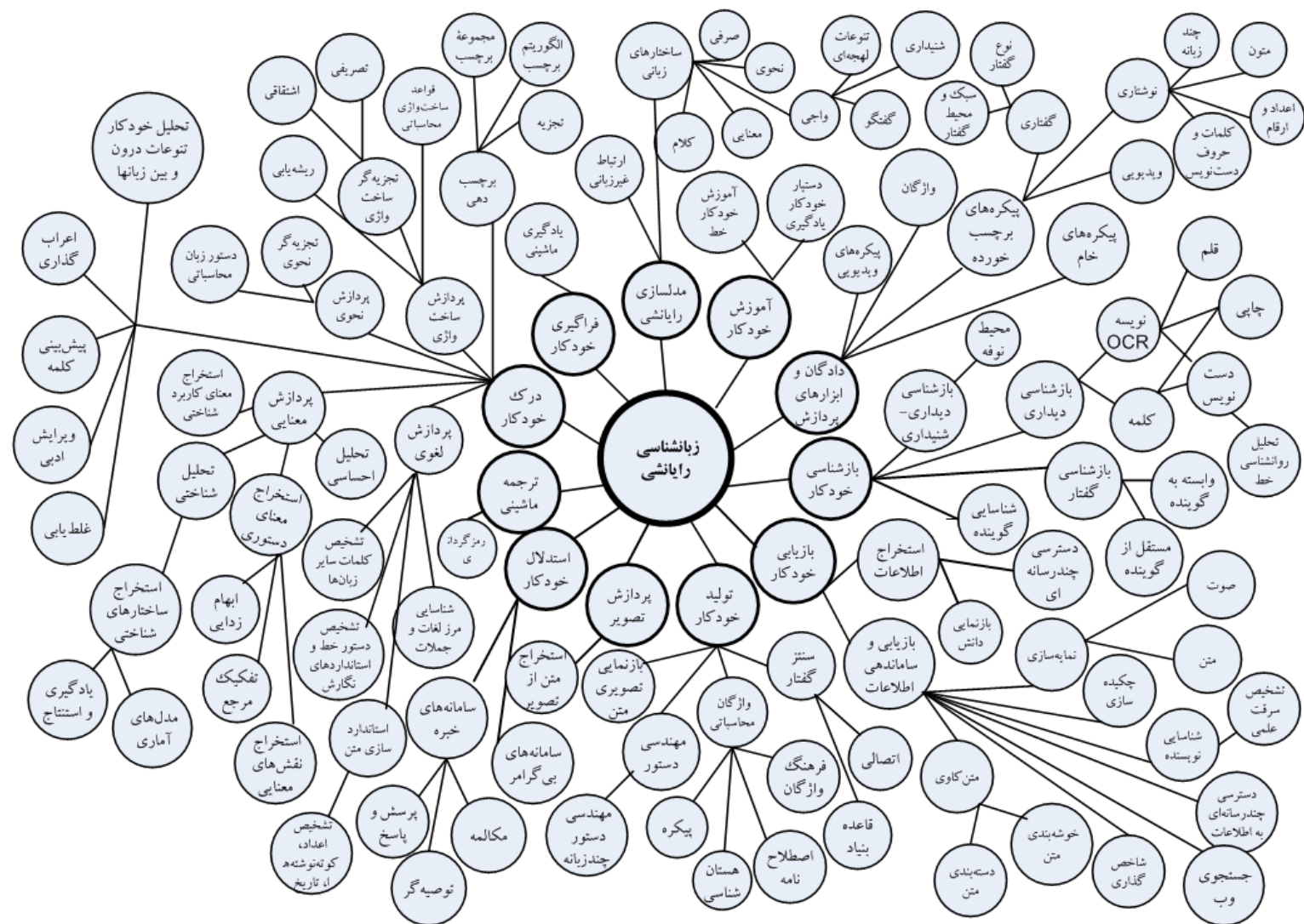


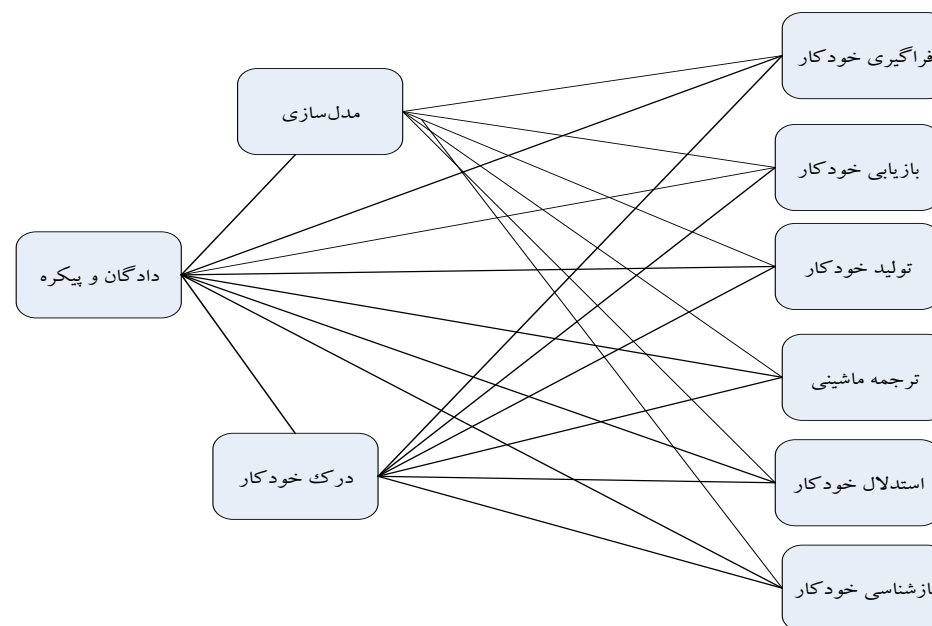


شکل ۱. زبان‌شناسی رایانشی حوزه‌ای میان‌رشته‌ای است (اقتباس از بولشاکوف و گلباخ ۲۰۰۴، ۱۸)

هدف این است که کامپیوترها قابلیت انجام فعالیت‌های مربوط به زبان انسان را بدست آورند. زبان انسان به دو صورت گفتار و نوشتار روی می‌دهد. فناوری‌های گفتار و متن براساس این دو نمود، زبان را تولید یا پردازش می‌کنند. در ارتباط، ما زبان را با سایر مدل‌های ارتباطی و سایر ابزارهای ارتباطی به هم می‌آمیزیم. ما گفتار را با ژست و حالات چهره ترکیب می‌کنیم. هدف فناوری‌های زبان ابداع مدل‌هایی است که توانایی تقلید همه این ابزارهای ارتباطی انسان را داشته باشند.

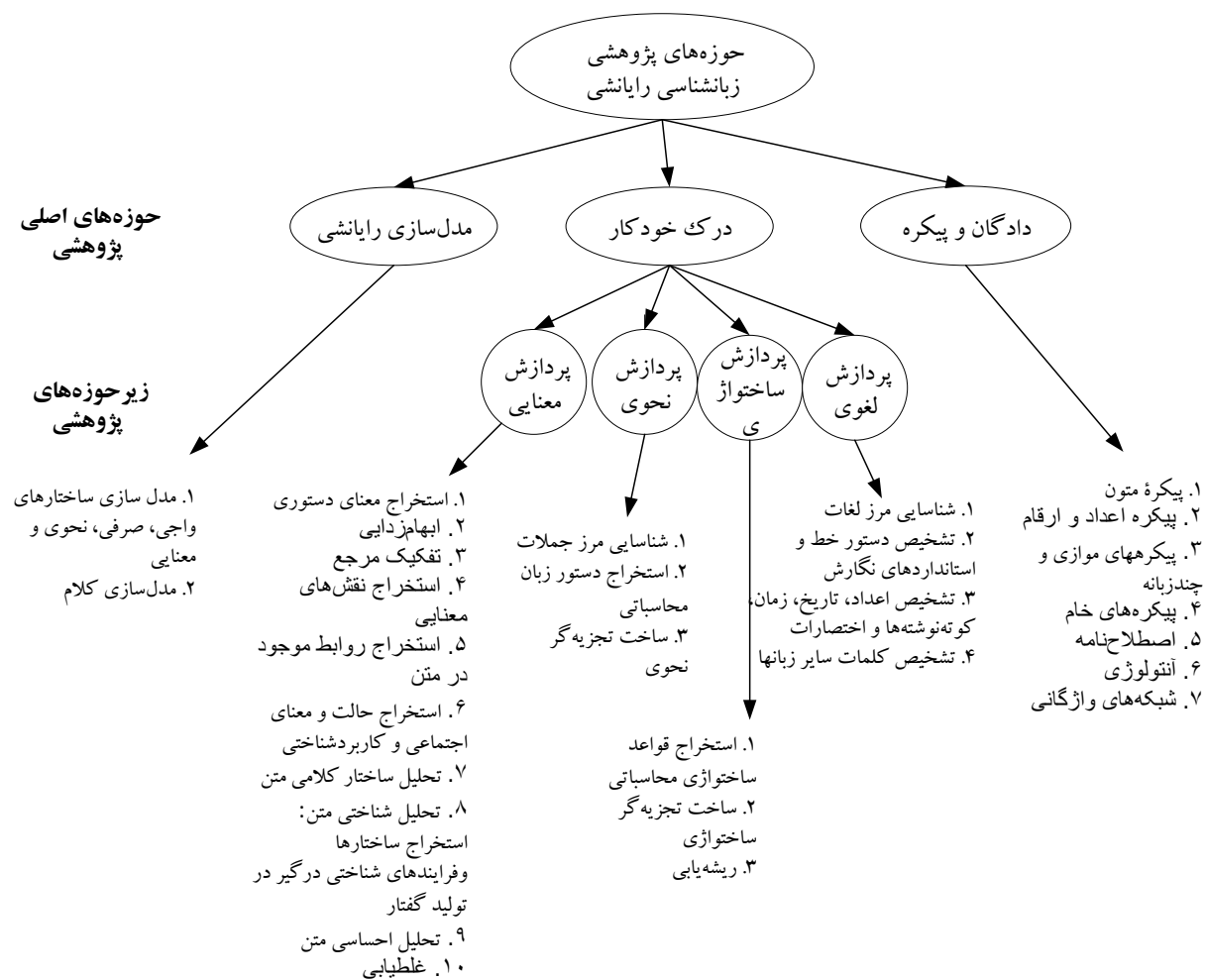
اصطلاح‌نامهٔ زبان‌شناسی سورگل و همکاران (۲۰۰۵، ۳۰-۳۲) مجموعه مقالات کنفرانس‌های زبان فارسی و رایانه (بی‌جن‌خان و صامتی ۱۳۸۹)، سومین همایش زبان‌شناسی رایانشی (۱۳۹۳)، جورافسکی و جیمز مارتین (۲۰۰۰) و نیز براساس برنامهٔ درسی و اطلاعات موجود در وب‌سایت‌های گروه‌های زبان‌شناسی رایانشی و پردازش زبان طبیعی در دانشگاه‌های استنفورد، واشنگتن، برکلی و غیره ۱۱ حوزهٔ اصلی و زیرحوزه‌های این حوزه‌های اصلی پژوهشی بدست آمد و با روش اصطلاحنامه‌ای به صورت یک مدل مفهومی به شکل زیر ترسیم شد:



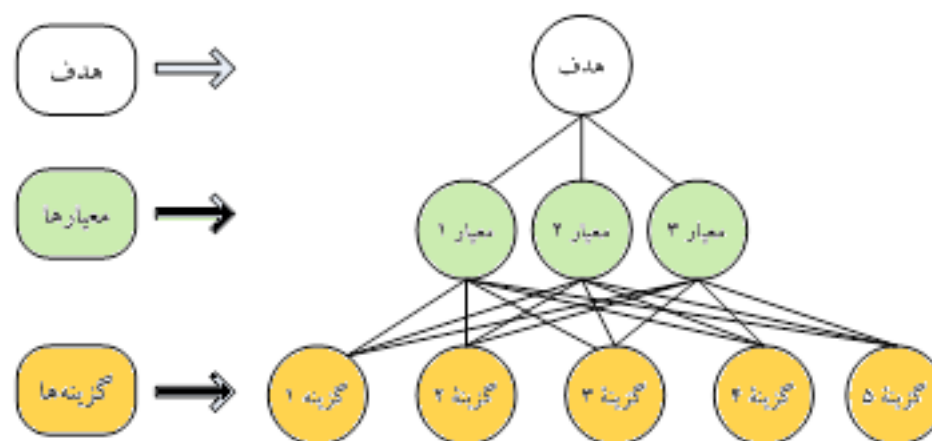


شکل ۲. شبکه پروژه حاصل از مدل مفهومی زبانشناسی رایانشی

براساس شبکه پروژه‌ها اولویت‌بندی در سه حوزه اصلی دادگان و پیکره، درک خودکار و مدل‌سازی رایانشی صورت گرفت. تعداد زیرحوزه‌های این سه حوزه اصلی ۲۹ مورد است که به عنوان گزینه‌ها یا کاندیداهای مقایسه وارد AHP شدند.

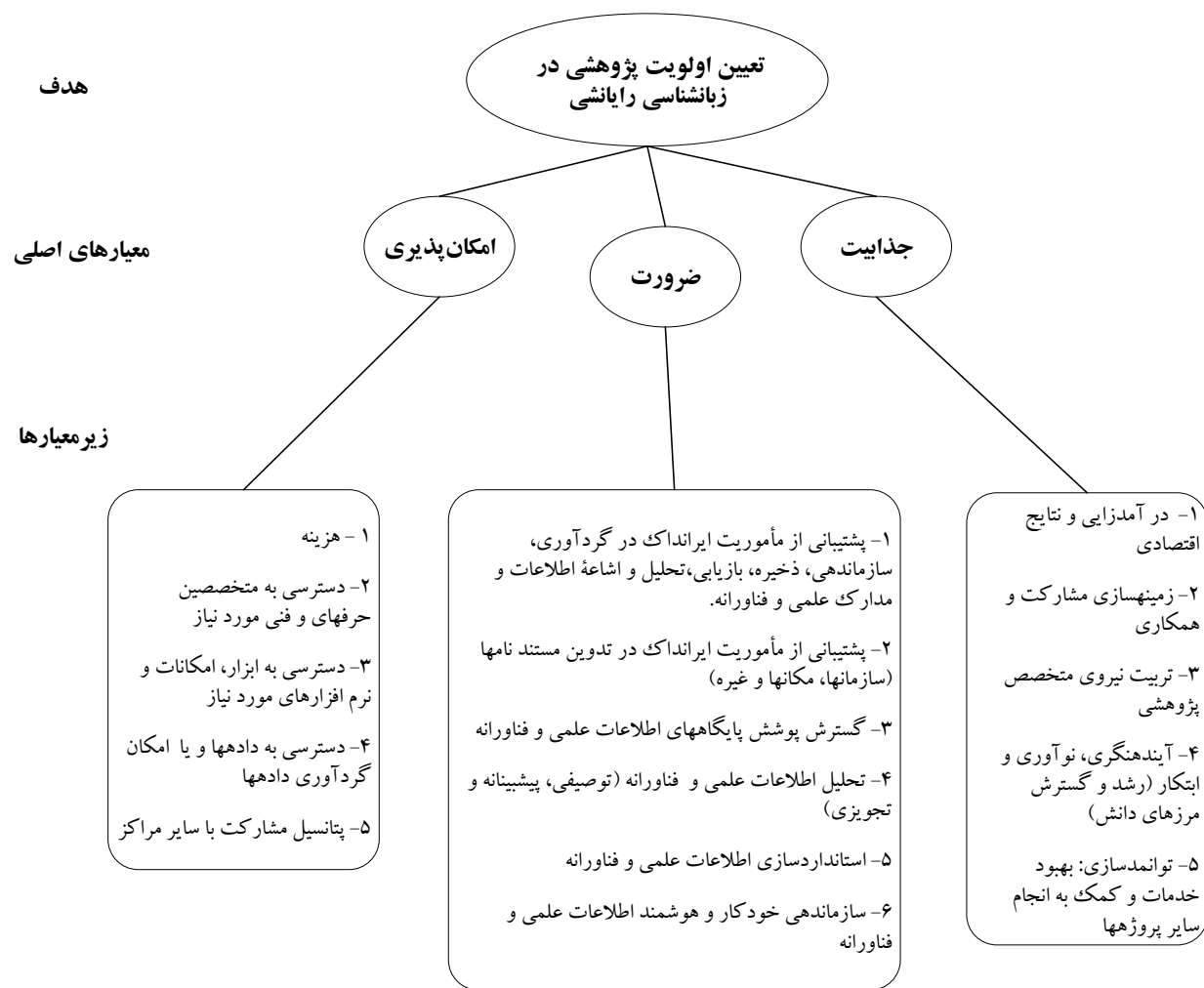


شکل ۶. فهرست ۲۹ حوزه پژوهشی که به عنوان گزینه‌های AHP تعیین شدند.



معیارهای اولیه: اسناد بالادستی پژوهشگاه علوم و فناوری اطلاعات ایران، شامل اساسنامه و برنامه استراتژیک ۱۳۸۸، سیاست‌های کلی علم و فناوری ۱۳۹۳ و نقشه جامع علمی کشور ۱۳۶۳.

فهرست نهایی معیارها: برگزاری جلسه‌های متعدد و نشست‌های بارش افکار با حضور خبرگان



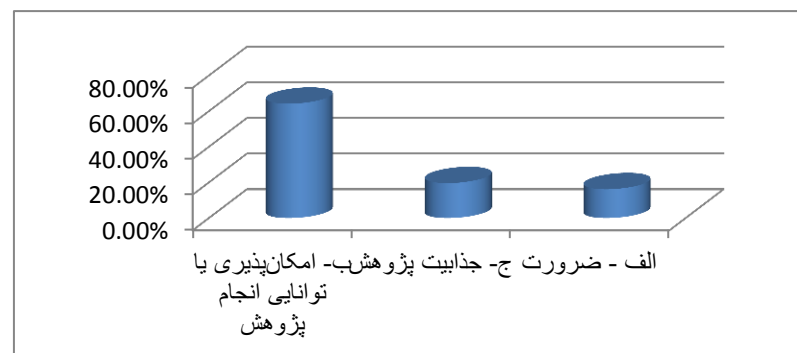
شکل ۷. فهرست معیارهای نهایی در مدل تصمیم

۲. تعیین وزن معیارها

برای مقایسه معیارها با یکدیگر و درنهایت وزن‌دهی آنها از پرسش‌نامه خبره و مقیاس نه درجه ساعتی (۱۹۸۰) استفاده شد. شیوه مقایسه زوجی معیارها و امتیازدهی آنها در جلسات گروهی و یا دونفره برای هر یک از اعضای هیئت علمی شرح داده شد.

سطح ۱	ضرورت	امکان‌پذیری با توانایی انجام پژوهش	جداپذیر پژوهش	جمع		
ضرورت	1	1/7	1/5	1 1/3		
امکان‌پذیری با توانایی انجام پژوهش	7	1	3	11		
جداپذیر پژوهش	5	1/3	1	6 1/3		
جمع	13	1 1/2	4 1/5	18 2/3		
سطح ۲ - الف ضرورت	پشتیبانی از مأموریت ایرانشادک در گردشگری، سازماندهی، ذخیره، بازاریابی، تحلیل و اشاعه اطلاعات و مدارک علمی و فناوری	پشتیبانی از مأموریت ایرانشادک در تدوین مستند نامها (سازمانها)، مکانها و غیره)	گسترش پوشش پایگاههای اطلاعات علمی و فناوری	تحلیل اطلاعات علمی و فناوری (توصیفی، پیش‌بینانه و تجویزی)	استانداردهای اطلاعات علمی و فناوری	سازماندهی خودکار و روشنشدن اطلاعات علمی و فناوری
پشتیبانی از مأموریت ایرانشادک در گردشگری، سازماندهی، ذخیره، بازاریابی، تحلیل و اشاعه اطلاعات و مدارک علمی و فناوری	1	1	1	1/3	1	1
پشتیبانی از مأموریت ایرانشادک در تدوین مستند نامها (سازمانها)، مکانها و غیره)	1	1	1	1/3	1	1
گسترش پوشش پایگاههای اطلاعات علمی و فناوری	1	1	1	1/3	1	1
تحلیل اطلاعات علمی و فناوری (توصیفی، پیش‌بینانه و تجویزی)	3	3	3	1	3	3
استانداردهای اطلاعات علمی و فناوری	1	1	1	1/3	1	1
سازماندهی خودکار و روشنشدن اطلاعات علمی و فناوری	1	1	1	1/3	1	1
جمع	8	8	8	2 2/3	8	8
سطح ۲ - ب امکان‌پذیری با توانایی انجام پژوهش	هزینه	دسترسی به متخصصین حرفه‌ای و فنی مورد نیاز	دسترسی به ابزار، امکانات و نرم افزارهای مورد نیاز	دسترسی به داده‌ها و یا امکان گردشگری داده‌ها	پتانسیل مشارکت با سایر مراکز	جمع
هزینه	1	1/3	3	3	5	12 1/3
دسترسی به متخصصین حرفه‌ای و فنی مورد نیاز	3	1	5	5	7	21
دسترسی به ابزار، امکانات و نرم افزارهای مورد نیاز	1/3	1/5	1	1	3	5 1/2
دسترسی به داده‌ها و یا امکان گردشگری داده‌ها	1/3	1/5	1	1	3	5 1/2
پتانسیل مشارکت با سایر مراکز	1/5	1/7	1/3	1/3	1	2
جمع	4 6/7	1 7/8	10 1/3	10 1/3	19	46 2/5
سطح ۲ - ج جداپذیر پژوهش	در آشنایی و نتایج اقتصادی	زمین‌سازی مشارکت و همکاری	تربیت نیروی متخصص پژوهشی	آپدیت‌نگری، نوآوری و ابتکار (رشد و گسترش مرزهای دانش)	بهبود خدمات و کمک به انجام سایر پروژه‌ها	جمع
در آشنایی و نتایج اقتصادی	1	3	1/3	1/3	1/3	5
زمین‌سازی مشارکت و همکاری	1/3	1	1/5	1/5	1/5	2
تربیت نیروی متخصص پژوهشی	3	5	1	1	1	11
آپدیت‌نگری، نوآوری و ابتکار (رشد و گسترش مرزهای دانش)	3	5	1	1	1	11
بهبود خدمات و کمک به انجام سایر پروژه‌ها	3	5	1	1	1	11
جمع	10 1/3	19	3 1/2	3 1/2	3 1/2	40

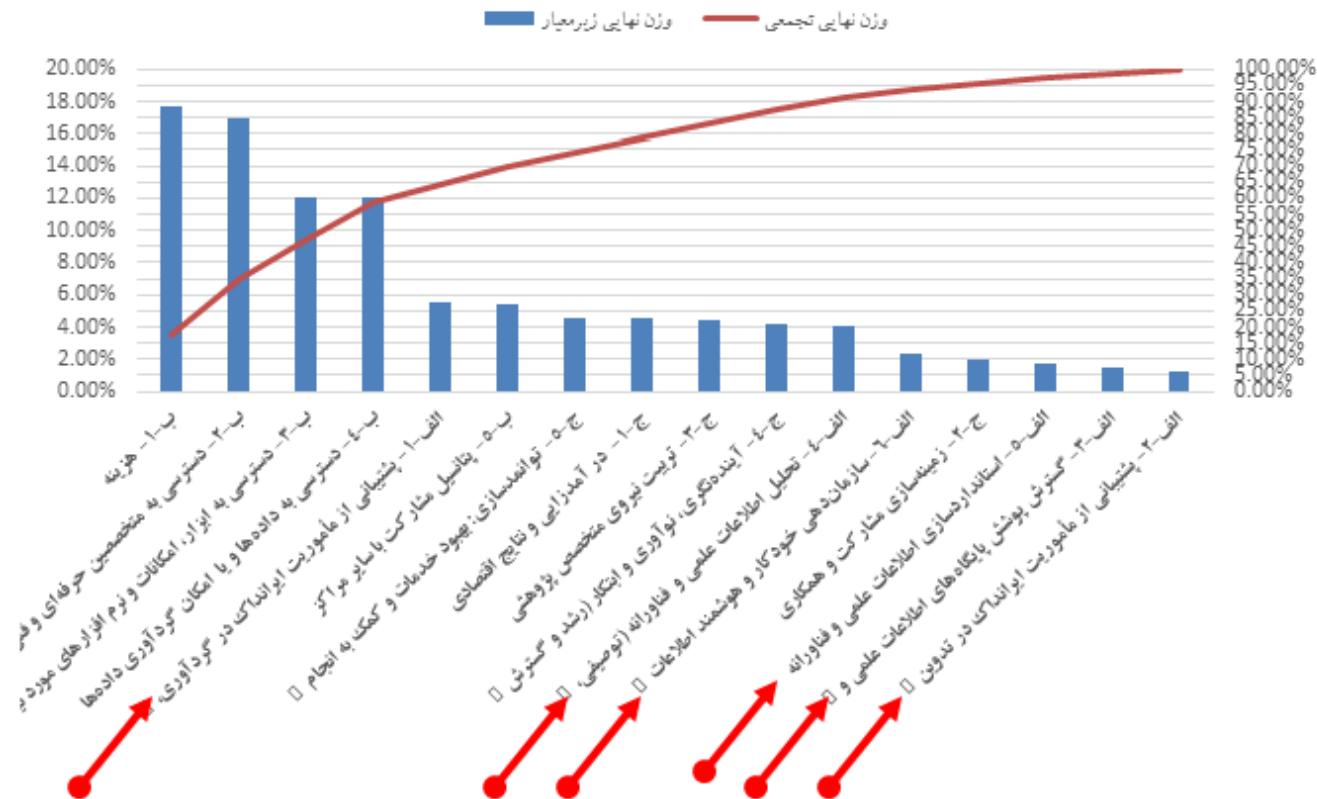
وزن نهایی معیارها (جدول ۱) نشان می‌دهد که معیار امکان‌پذیری یا توانایی انجام پژوهش به طور معناداری مهم‌تر و دارای وزن بسیار بالاتری نسبت به دو معیار دیگر است. نتایج نشان می‌دهد که معیار «امکان‌پذیری یا توانایی انجام پژوهش» با ۶۴ درصد امتیاز بیشترین وزن را برای کارشناسان زبانشناسی در پژوهشگاه دارد. این معیار به طور معناداری وزن بیشتری نسبت به دو معیار دیگر یعنی «جذابیت» با وزن تقریباً ۲۰ درصد و «ضرورت» با وزن ۱۶ درصد دارد. نمودار زیر وزن نهایی معیارهای اصلی و اولویت‌بندی آنها را نشان می‌دهد.



نمودار ۱. سلسله‌مراتب معیارهای اصلی پژوهش براساس AHP

براساس مقادیر بدست آمده در جدول ۱، نمودار ۲ رسم شد. این نمودار سلسله‌مراتب زیرمعیارهای پژوهشی که برای انتخاب حوزه‌های پژوهشی در گروه زبانشناسی رایانشی بکار گرفته شدند، را نشان می‌دهد.

وزن نهایی زیر معیارها



تحلیل سلسله‌مراتبی معیارها نشان می‌دهد که معیار ضرورت که بر اساس اسناد بالادستی پژوهشگاه بدست آمده، وزن بسیار کمتری نسبت به معیارهای امکان‌پذیری و جذابیت دارد.

۳ امتیاز حوزه‌های پژوهشی

در تحلیل سلسله‌مراتبی یا AHP، مقایسه زوجی گزینه‌ها براساس هریک از معیارها صورت می‌گیرد. اما اگر تعداد گزینه‌ها زیاد باشد، شدت نسبی آنها برحسب هر معیار تعیین می‌شود و امتیازدهی به صورت مقایسه زوجی نیست بلکه با تعیین درجات شدت گزینه‌ها نسبت به یک معیار صورت می‌گیرد و در واقع مقایسه نسبی گزینه‌ها برحسب هر معیار صورت می‌گیرد (Forman and Selly 2002, 59-61). فرمان و سلی^۲ (۲۰۰۲، ۵۹-۶۱) از این روش با عنوان رویکرد درجه‌بندی^۳ نام می‌برند و این روش را در مواجهه با تعداد زیاد گزینه‌ها یا کاندیداها مطرح می‌کنند. در این روش فهرست گزینه‌ها تهیه می‌شود و سپس درجه یا شدت هر یک از آنها نسبت به هر معیار ارزیابی می‌شود. فرمان و سلی (همان) ذکر می‌کنند که در روش درجه‌بندی، معیارها به صورت زوجی با یکدیگر مقایسه می‌شوند. اما به دلیل تعداد زیاد گزینه‌ها شدت یا درجه هر گزینه نسبت به هر معیار سنجیده می‌شود.

^۲ . Forman E., and Selly M. A.

^۳ . rating approach

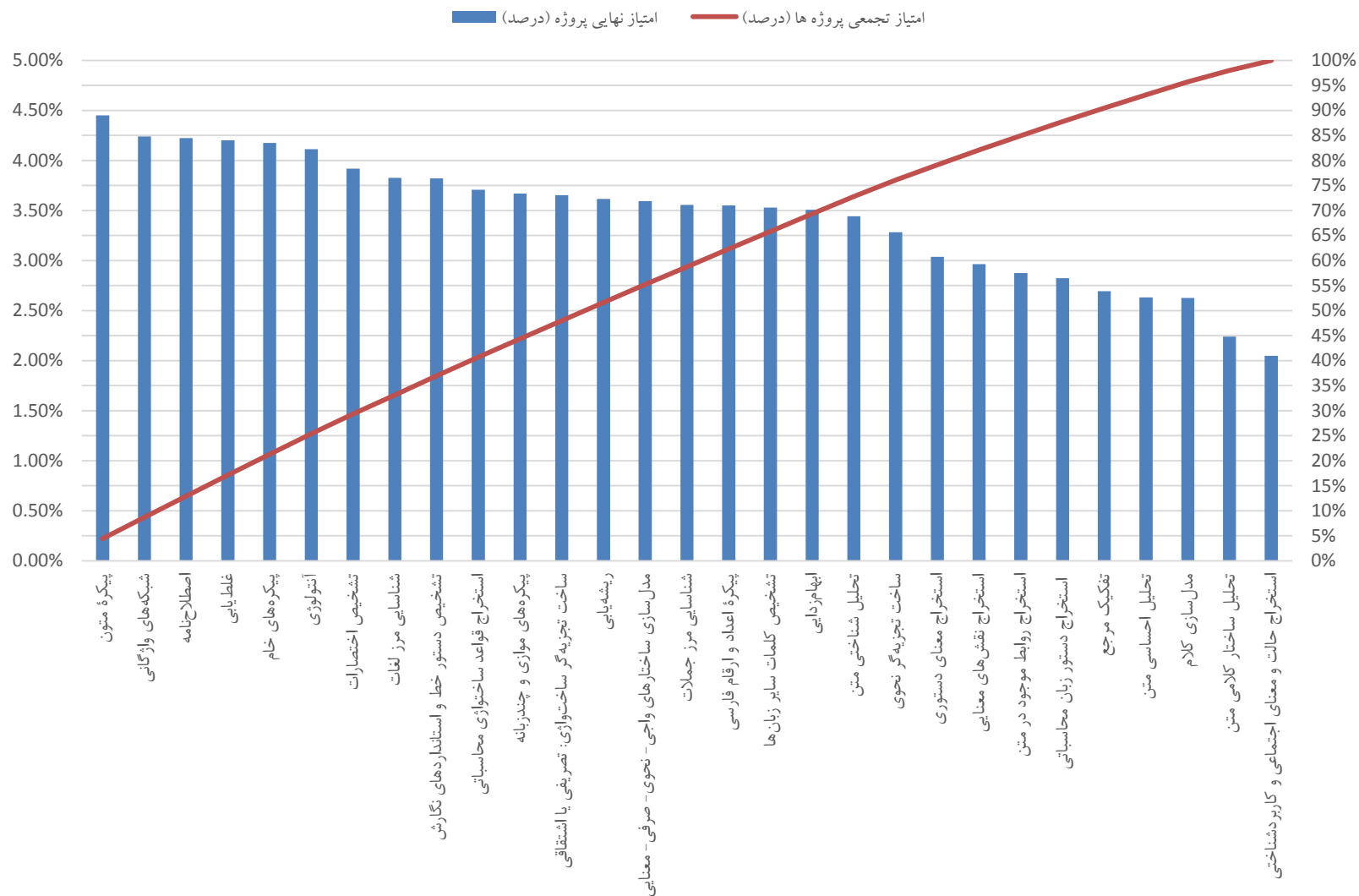
رتبه بندی آخرین سطح معیار																عناوین پروژه‌ها در حوزه زبانشناسی رایانشی	
۱ کمترین و ۷ بیشترین																	
16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1		
																در کت خودکار automated understanding	
																الف- پردازش نحوی	
6	7	7	5	4	5	6	7	7	4	7	7	7	7	7	7	۱. شناسایی مرز لغات	
5	7	7	6	5	6	7	6	7	3	6	7	6	7	7	7	۲. تشخیص دستور خط و استانداردهای نگارش	
6	7	7	5	4	5	7	4	7	2	5	7	5	6	6	7	۳. تشخیص اعداد، تاریخ، زمان، کلمات شش‌ه‌ا و انحصارات (text normalization)	
5	7	7	5	4	5	6	5	7	5	4	7	4	7	5	6	۴. تشخیص کلمات سایر زبان‌ها	
																ب- پردازش ساختاری	
6	7	7	6	5	6	7	4	7	4	7	7	7	6	7	7	۵. استخراج قواعد ساختاری معنایی	
5	7	7	5	6	5	7	7	7	5	7	7	7	5	7	7	۶. ساخت تجزیه گر ساخت وازی؛ تصریفی یا اشتقاقی (morphological parsing)	
6	7	7	6	5	6	7	7	7	3	7	7	7	7	7	7	۷. ریشه‌یابی	
																ج - پردازش نحوی	
5	7	7	6	5	6	7	7	7	2	6	7	7	7	7	7	۸. شناسایی مرز جملات	
6	7	7	5	6	6	6	6	7	2	6	7	5	5	6	6	۹. استخراج دستور زبان معنایی	
5	7	7	6	7	7	7	5	7	2	5	7	6	4	7	7	۱۰. ساخت تجزیه گر نحوی (syntactic parsing)	
																د- پردازش معنایی	
5	6	7	7	6	6	7	5	7	4	7	6	7	7	6	6	۱۱. استخراج معنای دستوری grammatical meaning	
6	7	7	6	5	6	7	4	6	5	4	6	7	7	7	7	۱۲. بهام‌زدایی word sense disambiguation	
6	7	7	6	4	5	6	5	7	4	6	7	6	5	5	6	۱۳. تکنیک مرجع Coreference (resolution)	
6	7	7	7	5	6	7	6	6	4	7	7	7	7	7	7	۱۴. استخراج نقش‌های معنایی semantic role labeling	
5	7	7	6	4	6	7	6	7	5	7	7	7	7	7	7	۱۵. استخراج روابط موجود (اختیارات گنی، کل به جز و ...) در متن Relation extraction	
6	7	7	6	5	5	7	5	6	4	7	7	7	6	6	5	۱۶. استخراج حالت و معنای اجسامی و کاربرد شناختی	
5	7	7	6	4	6	7	6	6	5	7	7	7	6	6	7	۱۷. تحلیل ساختار کلامی متن Discourse analysis	
6	7	7	6	4	5	5	6	6	4	7	7	6	5	7	7	۱۸. تحلیل شناختی متن: استخراج ساختارها و فرایندهای شناختی درگیر در تولید گفتار	
5	7	7	5	4	5	6	5	5	4	7	7	4	4	4	5	۱۹. تحلیل احساسی متن sentiment analysis (opinion analysis)	
7	7	7	6	4	4	5	6	6	5	7	7	7	6	7	7	۲۰. غلط‌یابی spell checking	
																مدل‌سازی رایانشی computational modeling	
7	7	7	6	5	6	7	6	7	4	7	7	7	6	7	7	۲۱. مدل‌سازی ساختارهای وابعی - نحوی - صرفی - معنایی	
6	7	7	6	4	5	6	5	6	4	7	7	5	4	6	6	۲۲. مدل‌سازی کلام discourse modeling	
																داده‌گان و پیکره database and corpus	
7	7	7	6	5	6	7	7	7	5	7	7	7	6	7	7	۲۳. پیکره متنون text corpus	
5	7	7	7	5	5	6	5	6	4	7	7	7	6	5	6	۲۴. پیکره اعداد و ارقام فارسی	
7	7	7	5	6	6	7	7	7	5	7	7	7	6	7	7	۲۵. پیکره‌های موازی و چندزبانه (multilingual corpora and Parallel corpora)	
7	7	7	5	6	6	7	7	7	6	7	7	7	6	5	6	۲۶. پیکره‌های خام	
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	۲۷. اصطلاح‌نامه	
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	۲۸. آنونوژی	
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	۲۹. شبکه‌های واژگانی lexical network	
172	202	203	172	148	166	192	170	193	125	189	201	187	176	186	192	جمع	

نتایج

سلسله مراتب حوزه های اصلی پژوهشی برای زبانشناسی رایانشی براساس AHP

حوزه های اصلی زبانشناسی رایانشی امتیاز نهایی پروژه (درصد)		
۱. درک خودکار		65.37%
○ پردازش معنایی		29.64%
○ پردازش لغوی		15.09%
○ پردازش ساختوازی		10.98%
○ پردازش نحوی		9.66%
۲. داده گان و پیکره		28.42%
۳. مدل سازی رایانشی		6.22%

عناوین پروژه‌ها در حوزهٔ زبان‌شناسی رایانشی



نتایج نشان می‌دهد که حوزه پژوهشی درک خودکار با بیشترین امتیاز، اولویت اول پژوهشی برای پژوهشگران محسوب می‌شود. براساس تحلیل سلسله‌مراتبی حوزه‌های پژوهشی دادگان و پیکره و مدل‌سازی رایانشی به ترتیب اولویت‌های دوم و سوم پژوهشی در حوزه زبانشناسی رایانشی هستند. اما در عین حال بیشترین امتیاز در سه اولویت اول مربوط به حوزه دادگان و پیکره یعنی زیرحوزه‌های پیکره متون، شبکه‌های واژگانی و اصطلاح‌نامه است.